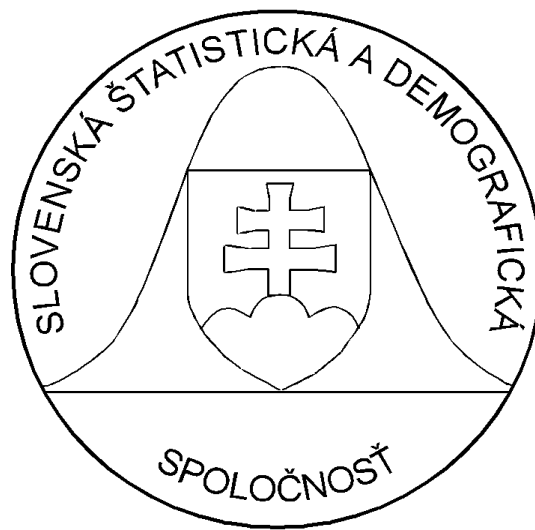


2/2011

FORUM STATISTICUM SLOVACUM



ISSN 1336-7420



9 771336 742001 20112 >



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



Naše najbližšie akcie:

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

EKOMSTAT 2011, 25. škola štatistiky

29. 5. - 3. 6. 2011, Trenčianske Teplice

Aplikácie metód na podporu rozhodovania vo vedeckej, technickej a spoločenskej praxi.

13. - 14. 9. 2011, STU Bratislava

FERNSTAT 2011

6. - 7. 10. 2011, UMB Banská Bystrica

13. Slovenská demografická konferencia "Mesto a vidiek"

19. - 21. 10. 2011, Nitriansky kraj

20. Medzinárodný seminár Výpočtová štatistika

1. - 2. 12. 2011, Infostat Bratislava

Prehliadka prác mladých štatistikov a demografov

1. 12. 2011, Infostat Bratislava

Regionálne akcie

priebežne

Pokyny pre autorov

Jednotlivé čísla vedeckého recenzovaného 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, 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 oponentom.

Príspevky nie sú honorované, poplatok za uverejnenie akceptovaného príspevku je minimálne 25 €. Za každú stranu navyše je poplatok 5 €.

Štruktúra príspevku: (Pri písaní príspevku využite elektronicкую šablónu: <http://www.sds.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

Abstrakt: Text abstraktu v slovenskom (prípadne českom) jazyku, max. 10 riadkov (štýl normálny: Time New Roman 12).

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

Kľúčové slová: Kľúčové slová v slovenskom (prípadne českom) jazyku, max. 2 riadky (štýl normálny: Time New Roman 12).

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

JEL classification: Uviesť kódy klasifikácie podľa pokynov v:

http://www.aeaweb.org/journal/jel_class_system.php

Vynechať riadok a nastaviť si medzery odseku pre nadpisy takto: medzera pred 12 pt a po 3 pt. Nasleduje 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 jednotlivých 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 a odsekmi nevynechávajúte. Nastavte si medzi odsekmi medzeru pred 0 pt a po 3 pt.

- 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): Uved'te svoju pracovnú adresu!!! (štýl **Nadpis 1: Time New Roman 12, bold, zarovnať vľavo, adresy vpísať do tabuľky bez orámovania s potrebným počtom stĺpcov a s 1 riadkom**):

Meno1 Priezvisko1, tituly1 (študenti ročník)

Pracovisko1 (študenti škola1)

Ulica1, 970 00 Mesto1

meno1.priezvisko1@mail.sk

Meno2 Priezvisko2, tituly2 (študenti ročník)

Pracovisko2 (študenti škola2)

Ulica2, 970 00 Mesto2

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

Redakčná rada:

RNDr. Peter Mach – *predseda*

Doc. Ing. Jozef Chajdiak, CSc. – *šéfredaktor*

RNDr. Ján Luha, CSc. – *vedecký tajomník*

členovia:

Ing. František Bernadič

Doc. RNDr. Branislav Bleha, PhD.

Ing. Mikuláš Cár, CSc.

Ing. Ján Cuper

Ing. Anna Janusová

Doc. RNDr. PaedDr. Stanislav Katina, PhD.

Prof. RNDr. Jozef Komorník, DrSc.

RNDr. Samuel Koróny, PhD.

Doc. RNDr. Bohdan Linda, CSc.

Prof. RNDr. Jozef Mládek, DrSc.

Doc. RNDr. Oľga Nánásiová, CSc.

Doc. RNDr. Karol Pastor, CSc.

Mgr. Michaela Potančoková, PhD.

Prof. RNDr. Rastislav Potocký, CSc.

Doc. RNDr. Viliam Páleník, PhD.

Ing. Marek Radvanský

Ing. Iveta Stankovičová, PhD.

Prof. RNDr. Beata Stehlíková, CSc.

Prof. RNDr. Anna Tirpáková, CSc.

Prof. RNDr. Michal Tkáč, CSc.

Doc. Ing. Vladimír Úradníček, PhD.

Ing. Boris Vaňo

Ing. Mária Vojtková, PhD.

Prof. RNDr. Gejza Wimmer, DrSc.

Ročník:

VII.

Číslo:

2/2011

Cena výtlačku: 20 EUR

Ročné predplatné: 80 EUR

ÚVOD

Vážené kolegyně, vážení kolegovia,

druhé číslo siedmeho ročníka vedeckého recenzovaného časopisu Slovenskej štatistickej a demografickej spoločnosti je zostavené z príspevkov, ktoré sú obsahovo orientované v súlade s tematikou konferencie Nitrianske štatistické dni. Táto akcia sa uskutočnila v dňoch 12. a 13. mája 2011 v Nitre. Konferenciu organizovala Slovenská štatistická a demografická spoločnosť v spolupráci s Univerzitou Konštantína Filozofa v Nitre.

Akciu z poverenia Výboru SŠDS, zorganizovali programový a organizačný výbor. Programový výbor: Prof. RNDr. Anna Tirpáková, CSc. – predseda, doc. Ing. Jozef Chajdiak, CSc., doc. PaedDr. Jana Kubanová, CSc., doc. RNDr. Bohdan Linda, CSc., RNDr. Ján Luha, CSc., RNDr. Jitka Poměnková, PhD., prof. RNDr. Beáta Stehliková, CSc., prof. RNDr. Ondrej Šedivý, CSc. a doc. RNDr. Marta Vrábelová, CSc. a Organizačný výbor: PaedDr. Janka Melušová, PhD. – predseda, doc. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc., doc. RNDr. Dagmar Markechová, CSc., RNDr. Kitti Vidermanová, PhD., PaedDr. Eva Uhrinová a Mgr. Mária Kóšová.

Na príprave a zostavení tohto čísla FORUM STATISTICUM SLOVACUM participovali: Prof. RNDr. Anna Tirpáková, CSc., Prof. RNDr. Beáta Stehliková, CSc., doc. RNDr. Dagmar Markechová, CSc. a doc. RNDr. Marta Vrábelová, CSc., doc. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc. a PaedDr. Janka Melušová, PhD.

Recenziu príspevkov zabezpečili: Prof. RNDr. Anna Tirpáková, CSc., Prof. RNDr. Beáta Stehliková, CSc., doc. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc., Ing. Zuzana Poláková, PhD., doc. RNDr. Dagmar Markechová, CSc. a doc. RNDr. Marta Vrábelová, CSc.

Organizátori touto cestou ďakujú pozvaným prednášajúcim prof. RNDr. Radkovi Mesiarovi, DrSc. a prof. RNDr. Gejzovi Wimmerovi, DrSc. za hodnotné prednášky.

Plošný prieskum kontaminácie pôdy na Slovensku Area soil contamination survey in Slovakia

Miriám Andrejiová, Zuzana Kimáková

Abstract: The article deals with area soil contamination survey in Slovakia. The first section of the article uses the method of multiple comparisons. In the second part with the using the cluster analysis are described clusters with characteristics similar observed parameters.

Key words: soil contamination, multi-dimensional comparison methods, cluster analysis

Kľúčové slová: kontaminácia pôdy, metódy viacrozmerného porovnávania, zhluková analýza

JEL classification: C38, C39

1. Úvod

Pôda predstavuje základný, v prevažnej miere neobnoviteľný prírodný zdroj a je v súčasnosti vystavená mnohým hrozbám a procesom degradácie (erózia, úbytok organickej hmoty, kontaminácia, zástavba, zhutnenie, salinizácia, záplavy a zosuvy atď.). Medzi závažnú degradáciu pôdy patrí kontaminácia pôd ťažkými kovmi. K najnebezpečnejším prvkom patrí kadmium, olovo, ortuť, arzén, nikel, chróm, meď a zinok.

Stupňovanie procesov, ktoré menia vlastností pôd je v posledných desaťročiach veľmi intenzívne. Preto sa na sledovanie týchto zmien zaviedol na základe Uznesenia vlády SR č. 620, zo 7. septembra 1993 monitoring pôdy, ktorý je v súčasnej dobe jedným z desiatich čiastkových monitorovacích systémov.

Plošný prieskum kontaminácie pôd (PPKP) v Slovenskej republike sa vykonáva v päťročných cykloch a je realizovaný Odborom životného prostredia a ekologického poľnohospodárstva Ústredného kontrolného a skúšobného ústavu poľnohospodárskeho v Bratislave. Vo vzorkách poľnohospodárskych pôd sa stanovujú organické aj anorganické kontaminanty. Hlavným cieľom už ukončených troch cyklov PPKP bolo sledovanie kontaminujúcich látok v pôdach vo vybraných katastrálnych územiach, ktoré boli určené na základe zvýšeného obsahu kontaminujúcich látok v pôdach z predchádzajúcich výskumov.

Od roku 2006 prebieha IV. cyklus PPKP. Predmetom plnenia úlohy je naďalej sledovať obsahy kontaminujúcich látok v pôdach vo vybraných územiach.

2. Metódy viacrozmerného porovnávania

Úroveň kontaminácie pôdy ťažkými kovmi vo vybraných okresoch Slovenska sme charakterizovali premennými: *Cr* (chróm), *Ni* (nikel), *Cd* (kadmium), *As* (arzén), *Hg* (ortuť), *Pb* (olovo), *H* a *K*. Prvých šesť premenných predstavuje obsah príslušného parametra v pôdnej vzorke [mg/kg]. Premenná *H* stanovuje množstvo kontrolovaných honov v hektároch a premenná *K* predstavuje množstvo nadlimitných honov [v ha]. Pri hodnotení plošnej kontaminácie pôdy sme brali do úvahy 58 okresov, pričom namerané hodnoty jednotlivých premenných sú výsledkami PPKP 2007 (odberový rok 2006). Prvok zinok a meď sme v našej analýze neuvažovali v dôsledku toho, že tieto parametre boli v danom odberovom roku sledované len v desiatich z 58 okresov.

Naším cieľom bolo usporiadanie 58 okresov na základe stupňa plošnej kontaminácie pôdy ťažkými kovmi. Pri porovnávaní sme zvolili metódy: jednoduchú metódu poradi, bodovaciu metódu, metódu normovanej premennej a metódu vzdialenosti od fiktívneho objektu. Premenné *Cr, Ni, As, Cd, Hg, Pb* a *K* sme považovali za destimulujúce premenné (pozitívnym rastom je ich pokles) a premennú *H* za stimulujúcu premennú.

Metóda poradi

Pri metóde poradi sme okresu s najnižšou hodnotou sledovaného parametra *Cr* priradili najvyššiu hodnotu rovnajúcu sa počtu okresov (58); okresu s druhou najnižšou hodnotou *Cr* poradie 57 atď. a tento postup sme opakovali pre každú premennú (*Ni, As, Cd, H, Pb* a *K*). V prípade rovnakých hodnôt sme príslušným okresom priradili priemerné poradie. Pri premennej *H* sme určovali poradie opačne: najvyšší počet bodov sme priradili okresu s najvyššou hodnotou sledovaného parametra. Výsledné poradie sledovaných okresov sme určili na základe súčtu jednotlivých poradi d_i .

V skupine okresov s najnižšou kontamináciou pôdy ťažkými kovmi sa nachádza napríklad okres Trnava, Zlaté Moravce, Veľký Krtíš a Trebišov. Okres Ružomberok, Žiar nad Hronom, Čadca, Myjava, Nové Zámky patria k okresom s najvyššou kontamináciou pôdy.

Tabuľka 1: Priemerné poradie 5 najlepších a 5 najhorších okresov – metóda poradi

Okres	Cr	Ni	Cd	As	Hg	Pb	H	K	d_i	P.č.
Trnava	51	27	52,5	48	56	53	57	39	383,5	1
Zlaté Moravce	52	40	55	54	48	57	27	39	372	2
Veľký Krtíš	55	47	58	57	42	54	7	39	359	3
Trebišov	50	31	52,5	41	53,5	49	42	39	358	4
Rimavská Sobota	35,5	46	56,5	44,5	32,5	48	54	39	356	5
Nové Zámky	18	2	17,5	9	19	38,5	36	8	148	54
Myjava	1	1	38	35,5	11	34,5	8	11	140	55
Čadca	8	38	1	15	26	1	17	7	113	56
Žiar nad Hronom	5	56	17,5	1	5,5	2	11	14	112	57
Ružomberok	4	21	2	17	29,5	5	26	2	106,5	58

Bodovacia metóda

Pri bodovacej metóde sme zistené hodnoty premenných nahradili počtom bodov. Pre destimulujúce premenné sme maximálny počet 100 bodov priradili okresu s najnižšou hodnotou daného parametra. Ostatným okresom sme priradili počet bodov podľa vzorca

$$z_{ij} = \frac{x_{\min,j}}{x_{ij}} \cdot 100\% \text{ pre } i = 1, 2, \dots, m, j = 1, 2, \dots, k \quad (1)$$

kde $m = 58$ je počet okresov, $k = 8$ je počet premenných, b_j je počet bodov pre j -tu premennú v i -tom okrese, x_{ij} je hodnota danej premennej, ktorá príslúcha i -temu okresu a $x_{\min,j}$ je minimálna hodnota príslušnej premennej.

Pre stimulujúcu premennú *H* sme maximálny počet bodov priradili okresu s najvyššou hodnotou a ostatným okresom sme priradili počet bodov podľa vzťahu

$$z_{ij} = \frac{x_{ij}}{x_{\max,j}} \cdot 100\% \text{ pre } i = 1, 2, \dots, m, j = 1, 2, \dots, k \quad (2)$$

Na základe celkového bodového hodnotenia d_i stanovíme poradie jednotlivých okresov. Výsledky porovnávaní sú v tabuľke 2.

Tabuľka 2: Priemerné poradie 5 najlepších a 5 najhorších okresov – bodovacia metóda

Okres	Cr	Ni	Cd	As	Hg	Pb	H	K	d _i	P.č.
Veľký Krtíš	82,19	50,29	100,00	78,95	62,71	89,66	5,85	100,0	569,64	1
Trnava	65,22	24,93	70,00	42,86	88,10	85,12	88,72	100,0	564,93	2
Rimavská Sobota	50,85	47,57	87,50	41,67	55,22	79,55	52,66	100,0	515,03	3
Krupina	72,29	95,65	50,00	51,72	68,52	48,56	25,46	100,0	512,20	4
Zlaté Moravce	67,42	33,59	77,78	51,72	71,15	92,71	17,42	100,0	511,78	5
Myjava	8,61	4,97	53,85	37,50	34,58	65,75	6,09	28,46	239,80	54
Žiar nad Hronom	18,18	88,00	41,18	6,10	13,45	34,29	7,11	29,91	238,22	55
Gelnica	60,00	47,06	50,00	15,79	1,07	37,71	9,33	14,69	235,64	56
Čadca	25,53	32,59	21,21	21,74	51,39	31,53	10,74	19,81	214,54	57
Ružomberok	15,63	22,56	23,33	23,44	52,86	39,86	16,74	12,70	207,11	58

Metóda normovanej premennej

Hodnoty destimulujúcich premenných normujeme podľa vzťahu

$$z_{ij} = \frac{\bar{x}_j - x_{ij}}{s_j}, \quad (3)$$

kde \bar{x}_j je aritmetický priemer a s_j je smerodajná odchýlka j -tej premennej.

Hodnoty premennej H normujeme podľa vzťahu

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}. \quad (4)$$

Na základe aritmetického priemeru normovaných hodnôt sme určili poradie okresov v závislosti od kontaminácie pôdy ťažkými kovmi. Výsledky sú v tabuľke 3.

Tabuľka 3: Priemerné poradie 5 najlepších a 5 najhorších okresov – normovaná premenná

Okres	Cr	Ni	Cd	As	Hg	Pb	H	K	d _i	P.č.
Trnava	0,67	-0,02	1,04	0,64	0,25	1,12	2,92	0,421	0,88	1
Rimavská Sobota	0,44	0,68	1,44	0,62	0,19	0,95	1,25	0,421	0,75	2
Veľký Krtíš	0,84	0,72	1,64	1,01	0,21	1,24	-0,93	0,421	0,64	3
Zlaté Moravce	0,70	0,36	1,24	0,78	0,23	1,31	-0,39	0,421	0,58	4
Trebišov	0,63	0,05	1,04	0,59	0,24	1,01	0,30	0,421	0,54	5
Čadca	-0,62	0,33	-3,56	-0,15	0,18	-2,87	-0,70	-0,681	-1,01	54
Ružomberok	-1,96	-0,17	-2,96	-0,03	0,19	-1,55	-0,42	-1,451	-1,05	55
Žiar nad Hronom	-1,48	1,04	-0,36	-4,25	-0,27	-2,36	-0,87	-0,217	-1,09	56
Gelnica	0,60	0,67	0,24	-0,75	-7,33	-1,83	-0,77	-1,160	-1,29	57
Myjava	-4,79	-5,94	0,44	0,52	0,11	0,43	-0,92	-0,263	-1,30	58

Metóda vzdialenosti od fiktívneho objektu

V prvom rade sme zvolili fiktívny objekt, ktorého súradnice závisia od typu premennej. Pre destimulujúce premenné platí

$$z_{oj} = \min_i \{z_{ij}\}, \text{ pre } i = 1, 2, \dots, m, \quad (5)$$

kde z_{ij} sú normované hodnoty premenných..

Súradnica fiktívneho bodu pre stimulujúcu premennú H je daná vzťahom

$$z_{oj} = \max_i \{z_{ij}\}, \text{ pre } i = 1, 2, \dots, m. \quad (6)$$

Poradie okresov (tabuľka 4) sme určili pomocou priemerných euklidovských vzdialeností d_i od fiktívneho objektu

$$d_i = \sqrt{\frac{1}{k} \sum_{j=1}^k (z_{ij} - z_{oj})^2}. \quad (7)$$

Tabuľka 4: Priemerné poradie 5 najlepších a 5 najhorších okresov – metóda vzdialenosti od fiktívneho objektu

Okres	Cr	Ni	Cd	As	Hg	Pb	H	K	d_i	P.č.
Veľký Krtíš	31,72	44,37	27,08	27,63	56,85	16,85	19,11	47,01	2,16	1
Zlaté Moravce	30,12	39,66	23,07	25,25	57,08	17,46	14,70	47,01	2,09	2
Spišská Nová Ves	32,03	36,93	25,04	21,45	55,72	16,96	18,51	47,01	2,09	3
Nové Mesto nad Váhom	28,45	37,79	21,19	24,10	57,12	15,67	15,76	47,01	2,06	4
Malacky	32,03	41,74	17,67	19,56	57,59	14,10	17,23	47,01	2,06	5
Nitra	24,82	31,74	19,39	4,54	55,32	13,53	0,47	25,16	1,74	54
Myjava	0,00	0,00	16,02	22,76	55,26	10,86	19,01	38,09	1,67	55
Ružomberok	7,98	33,22	0,36	17,76	56,48	1,75	14,94	24,84	1,65	56
Gelnica	29,04	43,71	14,46	12,23	0,00	1,07	17,72	27,83	1,59	57
Levice	27,50	37,08	12,98	8,92	54,44	3,81	0,00	0,00	1,58	58

Usporiadanie 10 okresov s najvyššou kontamináciou pôdy s využitím všetkých štyroch metód je v nasledujúcej tabuľke 5.

Tabuľka 5: Výsledné usporiadanie 10 okresov s najväčšou kontamináciou pôdy

Metóda poradi		Bodovacia metóda		Normovaná premenná		Vzdialenosť od fik. objektu	
Okres	P.č.	Okres	P.č.	Okres	P.č.	Okres	P.č.
Zvolen	49	Martin	49	Dunajská Streda	49	Nové Zámky	49
Senica	50	Zvolen	50	Rožňava	50	Čadca	50
Dunajská Streda	51	Levoča	51	Dolný Kubín	51	Košice – okolie	51
Gelnica	52	Nové Zámky	52	Žilina	52	Žiar nad Hronom	52
Dolný Kubín	53	Dolný Kubín	53	Levice	53	Martin	53
Nové Zámky	54	Myjava	54	Čadca	54	Nitra	54
Myjava	55	Žiar nad Hronom	55	Ružomberok	55	Myjava	55
Čadca	56	Gelnica	56	Žiar nad Hronom	56	Ružomberok	56
Žiar nad Hronom	57	Čadca	57	Gelnica	57	Gelnica	57
Ružomberok	58	Ružomberok	58	Myjava	58	Levice	58

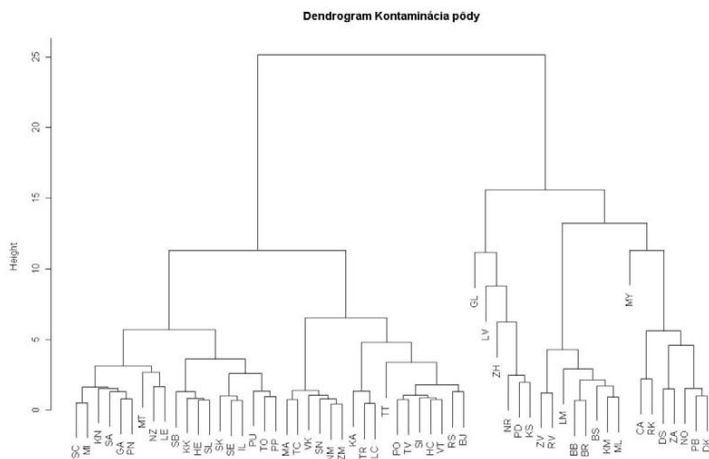
Okresy Ružomberok, Žiar nad Hronom, Čadca, Myjava a Gelnica sa vo všetkých použitých metódach objavujú na miestach s najväčšou kontamináciou pôdy ťažkými kovmi.

3. Zhuková analýza

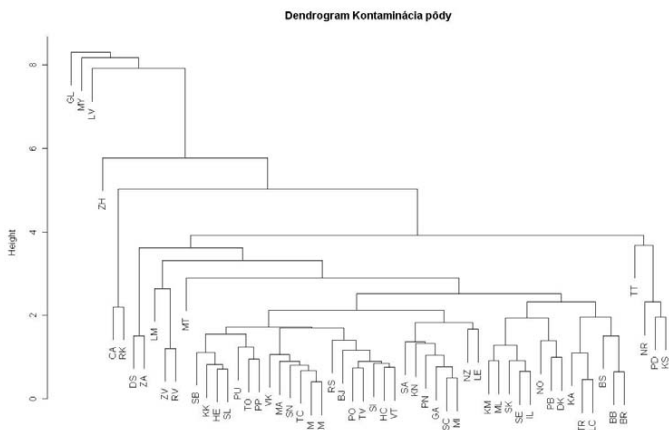
V ďalšej časti sme použili zhukovú analýzu, ktorá patrí medzi viacrozmerné štatistické metódy. Naším cieľom je rozdeliť, zoskupiť sledované okresy do zhukov s podobnými vlastnosťami. Aby sme odstránili vplyv jednotiek, v ktorých sú jednotlivé premenné merané, vstupné hodnoty najprv normalizujeme (využijeme výsledky získané pri metóde normovanej premennej).

V našej analýze boli postupne použité aglomeratívne hierarchické metódy: metóda najbližšieho suseda, metóda najvzdialenejšieho suseda, metóda priemernej väzby a Wardova metóda. Ako mieru vzdialenosti volíme Euklidovu vzdialenosť, ktorá je najčastejšie používanou mierou. Pred samotnou analýzou sme overili nekorelovanosť premenných.

Na obr.1 a obr.2 uvádzame dendrogram podobnosti okresov, ktorý sme získali Wardovou metódou a metódou priemernej väzby. Jeden rozmer reprezentujú okresy a druhý rozmer vzdialenosť medzi nimi.



Obrázok 1: Dendrogram – Wardova metóda



Obrázok 2: Dendrogram – metóda priemernej väzby

K overeniu správnosti výsledkov a k určeniu najlepšej metódy zhľukovania sme použili kofenetický korelačný koeficient CC. Najvyššia hodnota kofenetického korelačného koeficientu určuje najlepšiu metódu zhľukovania.

Tabuľka 6: Hodnoty kofenetického korelačného koeficientu

Metóda	CC	Metóda	CC
Metóda priemernej väzby	0,9412	Metóda najbližšieho suseda	0,9078
Wardova metóda	0,4693	Metóda najvzdialenejšieho suseda	0,7983

Z výsledkov vyplýva, že najlepšou metódou zhľukovania sa javí metóda priemernej väzby. Okresy Gelnica (GL), Myjava (MY), Levice (LV) a Žiar nad Hronom (ZH) vykazujú heterogénne vlastnosti a tvoria samostatnú skupinu (rovnako je to aj v prípade ostatných metód). Je to spôsobené práve tým, že v daných okresoch boli zistené v zvýšenej miere nadlimitné hodnoty sledovaných parametrov. Prvý zhľuk tvorí 46 okresov, v ktorých boli namerané priemerné hodnoty parametrov pod limitnými hodnotami. Tretí a štvrtý zhľuk tvoria opäť okresy s vyššími hodnotami parametrov, ktoré charakterizujú vyššiu kontamináciu pôdy.

Tabuľka 7: Zoradenie okresov do zhľukov – Metóda priemernej väzby

Zhľuk	Okresy		
Zhľuk 1	LM, ZV, RV, MT, SB, KK, HE, SL, PU, TO, PP, VK, MA, SN, TC, NM, ZM, RS, B, PO, TV, SI, HC, VT, SA, KN, PN, GA, SC, MI, NZ, LE, KM, ML, SK, SE, IL, NO, PB, DK, KA, TR, LC, BS, BB, BR		
Zhľuk 2	TT, NR, PD, KS		
Zhľuk 3	DS, ZA	Zhľuk 4	CA, RK
4 samostatné zhľuky	ZH, LV, MY, GL		

4. Záver

Prostredníctvom metód viacrozmerného porovnávania sme usporiadali okresy podľa stupňa kontaminácie pôdy ťažkými kovmi. Zhľukovou analýzou sme vytvorili skupiny vykazujúce určitú podobnosť z hľadiska kontaminácie pôdy. Okresy s najvyššou kontamináciou pôdy, v ktorých boli namerané nadlimitné hodnoty parametrov, tvoria samostatné zhľuky. Štatistické spracovanie bolo uskutočnené v programe R– package a EXCEL.

Príspevok bol spracovaný s podporou grantového projektu VEGA 1/0543/10.

5. Literatúra

- [1] PAŽITNÁ, M. – LABUDOVIČOVÁ, V. 2007. Metódy štatistického porovnávania. Bratislava: EKONÓM. 180s. ISBN 978-80-225-2285-4.
- [2] HEBÁK, P. . 2007. Vicerozmerné štatistické metódy (3). Praha: Infomatorium. 271 s. ISBN 80-7333-039-3.

Adresa autorov:

Miriám Andrejiová, RNDr. PhD.
KAMA SĽF, TU Košice
Letná 9
040 10 Košice
miriam.andrejiova@tuke.sk

Zuzana Kimáková, RNDr. PhD.
KAMA SĽF, TU Košice
Letná 9
040 10 Košice
zuzana.kimakova@tuke.sk

Regulačné diagramy a spôsobilosť výrobného procesu pomocou programu R

Quality control charts and process capability in program R

Miriám Andrejiová, Zuzana Kimáková

Abstract: Article deals with the production control charts and process capability analysis with the using free program R package.

Key words: control charts, process capability, capability indices, R package.

Kľúčové slová: regulačné diagramy, spôsobilosť procesu, indexy spôsobilosti procesu, program R.

JEL classification: C88, C19

1. Úvod

Regulačný diagram (tzv. Shewhartov diagram) je základným nástrojom štatistickej regulácie procesu. Je to grafický prostriedok, ktorý napomáha oddeliť náhodné variability procesu od systematických (vymedziteľných) príčin a umožňuje operatívne určovať, či je proces stabilný alebo nestabilný. Všeobecne existujú dva základné typy regulačných diagramov: *regulačné diagramy meraním* a *regulačné diagramy porovnávaním*.

Spôsobilosť procesu vyjadruje schopnosť výrobného procesu alebo zariadenia dosahovať predpísané požiadavky na kvalitu. Proces je spôsobilý, ak bol jeho rozptyl minimalizovaný tak, že spĺňa požiadavky zákazníka. Na zlepšenie procesu je potrebné proces navyše stabilizovať. Proces je stabilný, ak sa nemení jeho priemer ani priemerná odchýlka. Štatisticky zvládnutý proces znamená, že na proces pôsobia iba náhodné vplyvy a systematické (vymedziteľné) príčiny nestability sú eliminované.

K hodnoteniu spôsobilosti procesu sa používajú rôzne indexy spôsobilosti, ktoré porovnávajú predpísanú prípustnú variabilitu hodnôt danú tolerančnými medzami so skutočnou variabilitou sledovaného znaku kvality dosahovanou u štatisticky zvládnutého procesu. Obvykle sa používajú indexy C_p , C_{pl} , C_{pu} , C_{pk} a C_{pm} . Ich cieľom je posúdenie potenciálnej a skutočnej schopnosti procesu nepretržite poskytovať produkty, ktoré vyhovujú tolerančným medziam.

2. Program R

Program R patrí k tzv. programom s otvoreným zdrojovým kódom - Open Source Software a jeho najaktuálnejšiu verziu nájdeme na stránke <http://www.r-project.org/>.

Štatistické metódy riadenia kvality majú v tomto programe veľmi bohaté zastúpenie. K základnému balíčku „stat“, ktorý je implementovaný priamo v základnej ponuke programu, možno prostredníctvom archívu CRAN na domovskej stránke doinštalovať ďalšie balíčky, napr. „Hmisc“, „pastecs“ (popisná štatistika), „fBasics“, „nortest“ (testovanie normality), „outlier“ (testovanie odľahlých hodnôt), „lmtest“ (lineárna regresia) a iné. Štatistická regulácia procesu je zastúpená napr. v balíkoch „qcc“, „qAnalyst“, „IQCC“. Pre naše potreby budeme používať balík „qcc“.

Na vytvorenie základných typov regulačných diagramov meraním môžeme použiť príkaz `qcc()`, ktorý v tom najjednoduchšom prípade môže byť v tvare `qcc(data, type=" ")`. Pri výpočte regulačných hraníc sa zvyčajne uvažuje ich vzdialenosť 3σ od centrálnej priamky. Toto pravidlo je predvolené a zmenu môžeme uskutočniť pomocou príkazu `nsigmas`, napr. `qcc(data, type=" ", nsigmas =2)`.

Tabuľka 1: Druhy regulačných diagramov meraním – príkaz `qcc`

type	Regulačný diagram	type	Regulačný diagram
"xbar"	priemerov	"s"	smerodajných odchýlok
"R"	rozpätí	"xbar.one"	individuálnych hodnôt

Zmenou typu diagramov (`type="p"`, resp. `"np"`, `"c"`, `"u"`, `"g"`) môžeme vytvoriť aj regulačné diagramy porovnávaním.

3. Aplikácia v programe R

Na ilustráciu použijeme hodnoty, ktoré sme získali meraním výšky nitovania istej súčiastky [mm]. V programe EXCEL si vytvoríme napríklad súbor, v ktorom namerané hodnoty rozdelíme do 25 podskupín, pričom v každej podskupine je 5 hodnôt (obr.1). Sledovaný rozmer pomenujeme „Vyska“ a súbor uložíme na disk pod názvom `nitovanie.csv`.

	A	B	C
1	Vyska	Podskupina	
2	5,991	1	
3	5,92	1	
4	5,873	1	
5	5,862	1	
6	5,835	1	
7	6,082	2	
8	5,961	2	

Obrázok 1: Súbor `nitovanie.csv`

Analýzu procesu začneme načítaním pripraveného dátového súboru a príslušného balíka „`qcc`“, do programu R:

```
> library(qcc)
> table=read.table("nitovanie.csv", header=T, sep=";", dec=",")
```

Príkazom `qcc.groups()` vytvoríme podskupiny (obr.2):

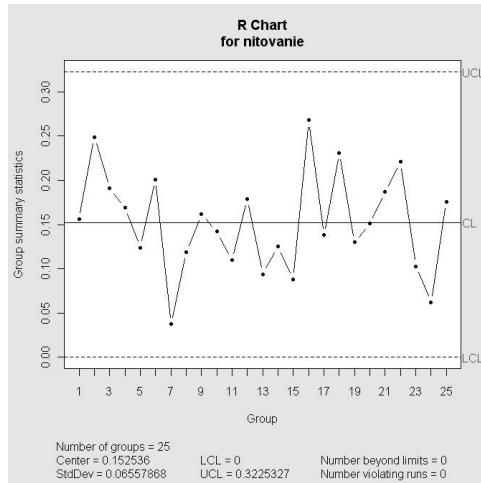
```
> nitovanie=qcc.groups(table$Vyska, table$Podskupina)
```

```
> nitovanie
      [,1] [,2] [,3] [,4] [,5]
1  5.991 5.9200 5.873 5.862 5.835
2  6.082 5.9610 6.108 5.833 6.015
3  6.035 5.9990 5.947 5.844 5.875
```

Obrázok 2: Vytvorenie podskupín – výstup z programu R

Príkaz na zostrojenie a vykreslenie regulačného diagramu rozpätí (R - diagram) je:

```
> objR=qcc(nitovanie, type="R")
```



Obrázok 3: Regulačný diagram (R – diagram)

Z výstupu vidíme, že centrálna priamka je $CL = 0,1525 = \bar{R}$, kde \bar{R} je priemerné výberové rozpätie. Dolná regulačná medza $LCL = 0$ a horná regulačná medza $UCL = 0,3225$. Z priebehu regulačného diagramu rozpätí (obr.3) vyplýva, že žiadna hodnota nie je mimo regulačných medzí LCL a UCL a môžeme predpokladať, že variabilita procesu je pod kontrolou. Pomocou príkazu `summary()` môžeme vytvoriť nielen grafické zobrazenie diagramov, ale aj celkový prehľad o jednotlivých regulačných hraniciach (obr.4).

```
> summary(objR)

Call:
qcc(data = nitovanie, type = "R")

R chart for nitovanie

Summary of group statistics:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0375 0.1190  0.1510  0.1536  0.1870  0.2750

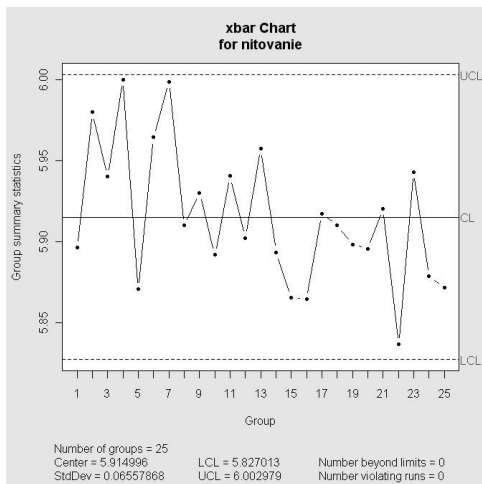
Group sample size: 5
Number of groups: 25
Center of group statistics: 0.153576
Standard deviation: 0.0660258

Control limits:
  LCL      UCL
  0 0.3247318
```

Obrázok 4: Príkaz `summary()`– výstup z programu R

V ďalšej časti pokračujeme zostrojením regulačného diagramu priemerov (x -diagram) a výpočtom príslušných regulačných medzí (obr. 5). Príkaz na zostrojenie x- diagramu:

```
> obj=qcc(nitovanie,type="xbar")
```

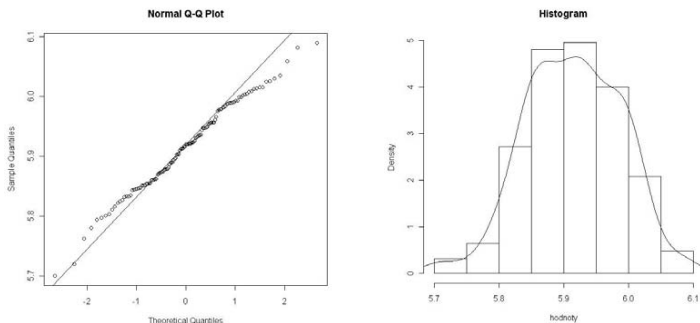


Obrázok 5: Regulačný diagram (\bar{x} – diagram)

Hodnota centrálnej priamky je $CL = 5,915 = \bar{\bar{x}}$, kde $\bar{\bar{x}}$ je celkový výberový priemer. Regulačný diagram ukazuje, že hodnoty ležia vo vnútri regulačných medzí $LCL = 5,827$ a $UCL = 6,0030$.

Podpredpokladom tvorby regulačných diagramov je podmienka normality nameraných hodnôt. To, či daný súbor hodnôt pochádza z normálneho rozdelenia, môžeme overiť graficky napríklad pomocou Q-Q grafu, histogramu, resp. pomocou testu normality (napr. Shapiro-Wilkov test). Príkazmi `qqnorm()` a `hist()` zostrojíme Q-Q graf a histogram početností (obr.6). Z ich priebehu a tvaru sa môže vizuálne posúdiť, či daný súbor pochádza z normálneho rozdelenia.

```
> qqnorm(table$Vyska); qqline(table$Vyska)
> hist(table$Vyska,breaks=7,prob=T,xlab="hodnoty",main="Histogram");
lines(density(table$Vyska))
```



Obrázok 6: Q-Q graf, histogram

Príkaz na Shapiro – Wilkov test normality, ktorý je implementovaný v základnom štatistickom balíku „stat“, je `shapiro.test()`.

Mierami spôsobilosti výrobného procesu sú indexy spôsobilosti. Na výpočet týchto indexov je nutné poznať smerodajnú odchýlku procesu, dolnú (LSL) a hornú (USL) tolerančnú medzu a cieľovú hodnotu (Target). Indexy spôsobilosti procesu môžeme vypočítať prostredníctvom funkcie `process.capability()`. Dolná tolerančná medza priemeru je $LSL=5,7$, horná tolerančná medza je $USL=6,1$, ktoré sú zadefinované príkazom `spec.limits()`:

```
> process.capability(obj,spec.limits=c(5.7,6.1))

Process Capability Analysis

Call:
process.capability(object = obj, spec.limits = c(5.7, 6.1))

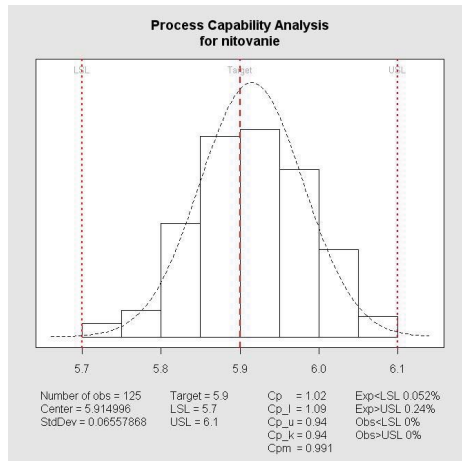
Number of obs = 125          Target = 5.9
Center = 5.914996          LSL = 5.7
StdDev = 0.06557868        USL = 6.1

Capability indices:

      Value    2.5%  97.5%
Cp    1.0166  0.8901  1.143
Cp_l  1.0928  0.9686  1.217
Cp_u  0.9404  0.8306  1.050
Cp_k  0.9404  0.8096  1.071
Cpm   0.9910  0.8652  1.117
```

Obrázok 7: Indexy spôsobilosti - výstup z programu R

Z výsledkov analýzy (obr.7) vyplýva, že proces je z pohľadu dodržiavania predpísaných hraníc nespôsobilý (hodnoty indexov C_{pu} , C_{pk} a C_{pm} sú menšie ako 1, index C_p je menší ako 1,33).

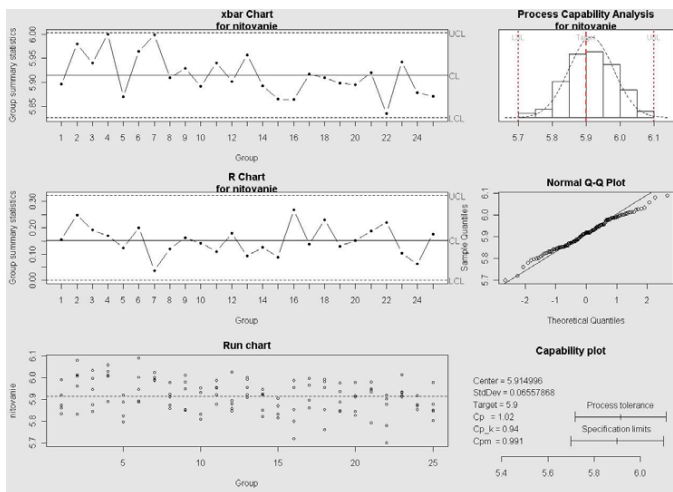


Obrázok 8: Grafická prezentácia spôsobilosti procesu

Pri určení indexu spôsobilosti C_{pm} sa použila cieľová hodnota $T=5,9$, ktorá predstavuje stred intervalu určeného hranicami USL a LSL . Funkciou „target=“ je možné danú hodnotu spresniť podľa požiadaviek. Grafickým výstupom je histogram spolu

s výpočtom daných indexov spôsobilosti (obr.8). Ďalší príkaz umožňuje vykonať záverečnú analýzu spôsobilosti procesu pomocou šiestich grafov (obr.9):

```
> process.capability.sixpack(obj, spec.limits=c(5.7, 6.1))
```



Obrázok 9: Záverečná analýza procesu

4. Záver

Balík „qcc“, ktorý sme sa snažili v článku priblížiť, umožňuje overiť aj testy vymedziteľných príčin (`shewhart.rules()`), vytvoriť Pareto diagram (`pareto.chart()`), diagram príčin a následkov (`cause.and.effect()`) a graf operatívnej charakteristiky (`oc.curves()`) a iné. K ďalším možnostiam patrí vytvorenie regulačných kumulatívnych súm (`cusum()`) alebo regulačný diagram exponenciálne vážených kĺzavých priemerov (`ewma()`). Podrobnejšie využitie balíka a ďalšie možnosti štatistického riadenia kvality v programe R je možné nájsť na oficiálnej stránke <http://www.r-project.org/>.

Príspevok bol spracovaný s podporou grantového projektu VEGA 1/0543/10.

5. Literatúra

- [1]TEREK, M. – HRNČIAROVÁ, E. 2004. Štatistické riadenie kvality. Bratislava: Iura Edition. 234 s. ISBN 80-89047-97-1.
- [2]SCRUCCA, L. 2010. Package „qcc“. In: <http://cran.r-project.org/web/packages/qcc/qcc.pdf>

Adresa autorov:

Miriám Andrejiová, RNDr. PhD.
KAMA SĽF, TU Košice
Letná 9
040 10 Košice
miriam.andrejiova@tuke.sk

Zuzana Kimáková, RNDr. PhD.
KAMA SĽF, TU Košice
Letná 9
040 10 Košice
zuzana.kimakova@tuke.sk

**Aplikácia metód viackriteriálneho hodnotenia
pri skúmaní vplyvu ekonomicko-právnych aspektov
na zakladanie nadnárodných obchodných spoločností v Európskej únii**
**An application of multicriterial assessment methods
in the investigation of the influence of economic-legislative factors
upon the founding of transnational corporations in the European Union**

Martin Balajty, Martin Boďa

Abstract: In the article, the selected member states of the European Union are assessed from the perspective of convenience of establishing headquarters for entrepreneurial activities. For this purpose, a number of standard methods for multicriteria methods of assessment are employed and the selected 22 member states are lined up according to their convenience for setting up an enterprise with respect to the chosen 10 criteria. The final rankings are evaluated statistically and it found that various methods are comparable in the results.

Key words: multicriterial assessment methods, the Saaty method, the SPAN method, Kendall's coefficient of concordance.

Kľúčové slová: metódy viackriteriálneho hodnotenia, Saatyho metóda, metóda SPAN, Kendallov koeficient konkordancie.

JEL classification: M13.

1. Úvod

Každý začínajúci alebo expandujúci podnik sa musí rozhodnúť pre umiestnenie svojho sídla, výrobnjej jednotky alebo iného typu afilácie. Menšie podniky svoje rozhodnutie o výbere vhodnej lokality obmedzujú na rádius neprekračujúci hranice štátu, väčšie podniky a osobitne transnacionálne korporácie majú možnosť výberu sídla v inom štáte s vhodnejšou skladbou ekonomických a legislatívnych faktorov ovplyvňujúcich ich podnikanie. V článku sú na základe vybraných rámcových ekonomických a právnych aspektov kľúčových pre činnosť podnikateľských subjektov hodnotené vybrané členské štáty Európskej únie z hľadiska ich atraktívnosti pre založenie obchodnej spoločnosti nadnárodnej formy. Inšpiráciou pre výber ukazovateľov bola štúdia Dvořáčka et al. (2006, s. 33-55) a pri hodnotení štátov Európskej únie ako variantov pre umiestnenie sídla obchodnej spoločnosti nadnárodnej formy boli uplatnené štandardné metódy viackriteriálneho hodnotenia variantov.

2. Dátová základňa a metódy

Hodnotenia sa zúčastňovalo 22 členských štátov Európskej únie: Belgicko, Bulharsko, Cyprus, Česká republika, Dánsko, Estónsko, Fínsko, Francúzsko, Litva, Lotyšsko, Luxembursko, Maďarsko, Nemecko, Poľsko, Portugalsko, Rakúsko, Rumunsko, Slovensko, Slovinsko, Veľká Británia, Španielsko a Švédsko. Pre ostatné štáty (Grécko, Írsko, Holandsko, Španielsko a Taliansko) neboli v čase analýzy dostupné všetky potrebné údaje.

Hodnotenie sa uskutočňovalo na základe desiatich indikátorov, ktorých obsahovú náplň a zdroj údajov približuje tabuľka č. 1. Indikátory boli vybrané v konzistencii s podobnou štúdiou Dvořáčka et al. (2006, s. 33-55) a pri získavaní údajov limitujúcou okolnosťou bola časová dostupnosť údajov a dostupnosť dát údajov. V niektorých prípadoch sa nemohli pre úplnú absenciu dát použiť aktuálnejšie údaje alebo museli byť volené k neskoršiemu roku vykazovania, aby sa veľmi nezúžila množina členských štátov Európskej únie zúčastňujúcich sa hodnotenia. Treba poznamenať, že vzhľadom na použité metódy merné jednotky

jednotlivých indikátorov nemali vplyv na výsledky analýzy a počas analýzy boli buď nahradené poradím, alebo ich vplyv bol anulovaný pomerovaním.

Dátová základňa predstavovala maticu X o 22 riadkoch a 10 stĺpcoch s typickým elementom x_{ij} , ktorý zapisuje hodnotu j -teho kritéria ($j \in \{1, \dots, 10\}$) zistenú na i -tom členskom štáte Európskej únie ($i \in \{1, \dots, 22\}$). Pre každý indikátor j ($j \in \{1, \dots, 10\}$) bol stanovený logický vplyv na podnikanie, pričom sa uvažovali iba dve možnosti: (+) – ak vyššia hodnota ukazovateľa stimuluje podnikanie, resp. (–) – ak vyššia hodnota ukazovateľa pôsobí vo všeobecnosti negatívne na podnikanie. Charakter ukazovateľa je tiež obsahom tabuľky č. 1. Každému indikátoru j ($j \in \{1, \dots, 10\}$) bola pridelená váha vyjadrujúca relatívny význam príslušného indikátora. Pri stanovovaní váh bola využitá Saatyho metóda a modifikovaná metóda SPAN (*Successive Proportionate Additive Numeration*).

Tabuľka 1. Uvažované ekonomické a právne faktory a ich údajový zdroj

Označenie	Obsah	Merné jednotky	Rok	Zdroj dát	Charakter
HDPpc	Hrubý domáci produkt na obyvateľa v parite kúpnej sily (vzhľadom k EÚ27)	Index	2008	Eurostat	+
rHDP	Medziročné tempo rastu hrubého domáceho produktu	%	2009	Eurostat	+
INFL	Priemerná medziročná percentuálna zmena inflácie meraná harmonizovaným indexom spotrebiteľských cien	%	2008	Eurostat	-
NEZ	Miera nezamestnanosti	%	2008	Eurostat	+
DZ	Celkove danové zatavenie (vrátane odvodového zatavenia) merané ako podiel na reálnom hrubom domácom produkte	%	2007	Eurostat	-
HDS	Hustota dopravnej siete meraná hustotou železničnej siete	km/ 1000 km ²	2005	Eurostat	+
HNP	Hodinové náklady práce	EUR	2007	Eurostat	-
PP	Produktivita práce na zamestnanca meraná hrubým domácim produktom v parite kúpnej sily (vzhľadom k EÚ27)	Index	2009	Eurostat	+
MMM	Minimálna mesačná mzda	EUR	2008	Eurostat	-
SS	Podiel populácie s prinajmenej stredoškolským vzdelaním (vzhľadom k populácii od 25 do 64 rokov)	%	2008	Eurostat	+

Zdroj: vlastné spracovanie.

Východiskom Saatyho metódy je tzv. Saatyho matica preferencií S , v tomto prípade o rozmeroch 10×10 (ukazovateľ \times ukazovateľ), v ktorej sú zaznamenané subjektívne preferencie medzi jednotlivými ukazovateľmi. Typický element matice s_{kl} vyjadruje, koľkokrát je k -ty riadkový indikátor ($k \in \{1, \dots, 10\}$) významnejší než l -tý stĺpcový indikátor ($l \in \{1, \dots, 10\}$). Na diagonále matice S sa nachádzajú jednotky, pretože ľubovoľný i -ty indikátor (v riadku) je rovnako významný ako on sám (v stĺpci), čiže $s_{jj} = 1$ pre $\forall j \in \{1, \dots, 10\}$ a mimodiagonálne elementy sú recipročne symetrické a platí $s_{kl} = 1/s_{lk}$ pre $\forall k, l \in \{1, \dots, 10\}$. Váhy jednotlivých ukazovateľov v_j pre $\forall j \in \{1, \dots, 10\}$ v štandardizovanom tvare sa získali riešením optimalizačnej úlohy.

$$\sum_{k=1}^{k=10} \sum_{l=1}^{l=10} \left(s_{kl} - \frac{v_k}{v_l} \right)^2 = \min \quad \text{za podmienok} \quad \begin{cases} v_1 + \dots + v_{10} = 1 \\ v_1, \dots, v_{10} > 0 \end{cases} \quad (1)$$

Saatyho matica so zvolenými párovými preferenciami a výsledné váhy je obsahom tabuľky č. 2.

Metóda SPAN spočíva v expertnom posúdení váhy jednotlivých indikátorov. Je stanovená skupina expertov a každému expertovi je pridelený rovnaký počet bodov (obvykle pre jednoduchosť 100 bodov). Každý expert rozdelí svoje body na dve časti. Prvá časť je určená ostatným expertom a druhá časť je určená pre ohodnotenie významnosti zvažovaných hodnotiacich indikátorov. Pomer medzi rozdelenými bodmi zostáva počas celého hodnotiaceho procesu konštantný. Každý expert rozdelí body vyhradené ostatným expertom

medzi jednotlivých expertov a body pridelené vyhradené jednotlivým indikátorom medzi jednotlivé indikátory. Určené pomery rozdelených bodov sú opäť pre celý proces nemenné. V uvedenom pridelovaní bodov sa iteruje v ďalších kolách až do okamihu, kedy sa stabilizuje celkový súčet bodov pridelený jednotlivým ukazovateľom a je s požadovanou presnosťou rovný celkovému počiatočnému počtu všetkých bodov. Následne sa získané hodnoty zoštandardizujú (Ficzová et al., 2002, s. 132). Expertnú skupinu tvorili interní pedagógovia Ekonomickej fakulty Univerzity Mateja Bela v Banskej Bystrici (Jaroslav Ďaďo, st., Lena Theodolides, Dagmar Kokavcová, Silvia Fraňová a Luboš Elexa) s dostatočnými znalosťami z oblasti medzinárodného podnikania a manažmentu, a preto by mali byť spôsobilí objektívne ohodnotiť jednotlivé aspekty. Výsledné váhy sú uvedené v tabuľke č. 2.

Tabuľka 2. Saatyho matica a pridelené váhy Saatyho metódou a metódou SPAN

	Saatyho matica s párovými preferenciami										Pridelené váhy	
	HDPpc	rHDP	INFL	NEZ	DZ	HDS	HNP	PP	MMM	SS	Saaty	SPAN
HDPpc	1	1/3	1/5	1	1/9	3	1/7	1/5	1/7	3	0.028	0.027
rHDP	3	1	1/3	1	1/9	1/3	1/5	1/7	1/5	1	0.030	0.065
INFL	5	3	1	3	1/5	3	1/5	1/5	1/3	3	0.059	0.062
NEZ	1	1	1/3	1	1/7	3	1/5	1/3	1	3	0.045	0.042
DZ	9	9	5	7	1	7	1	3	3	9	0.265	0.235
HDS	1/3	3	1/3	1/3	1/7	1	1/7	1/3	1/3	1	0.036	0.150
HNP	7	5	5	5	1	7	1	1	3	9	0.217	0.115
PP	5	7	5	3	1/3	3	1	1	1	5	0.155	0.149
MMM	7	5	3	1	1/3	3	1/3	1	1	5	0.137	0.110
SS	1/3	1	1/3	1/3	1/9	1	1/9	1/5	1/5	1	0.027	0.045

Zdroj: vlastné spracovanie.

Pri samotnom hodnotení sme aplikovali štyri metódy: metódu poradií, bodovaciu metódu, metódu normovanej premennej a metódu vzdialenosti od fiktívneho objektu.

Metóda poradií je najjednoduchšou metódou, ktorá spočíva v tom, že štáty v súbore sa zoradili podľa jednotlivých vybraných ukazovateľov. Štát, ktorý v danom ukazovateli dosahuje najlepšiu hodnotu, bolo priradené poradie 1, štátu s druhou najlepšou hodnotou ukazovateľa bolo priradené poradie 2 atď. až štátu s najhoršou hodnotou bolo priradená poradie 10. Integrovaný ukazovateľ metódy poradií sa definuje ako vážený súčet poradií štátov v jednotlivých ukazovateľoch. Štát s najnižšou hodnotou integrovaného ukazovateľa je vo výslednom poradí na prvom mieste, štát s druhou najnižšou hodnotou na druhom mieste atď. až štát s najvyššou hodnotou integrovaného ukazovateľa je hodnotený na poslednom, 22. mieste.

Pri bodovacej metóde pri každom ukazovateli bolo priradených štátov, ktorý dosiahol najlepšiu hodnotu daného ukazovateľa, 100 bodov a ostatným štátom bol priradený nižší počet bodov určený pomerom ich hodnoty pre daný ukazovateľ a najlepšej hodnoty pomocou vzťahu:

ak ukazovateľ má charakter (+)

$$b_{ij} = \frac{x_{ij}}{\max_i x_{ij}} \cdot 100$$

ak ukazovateľ má charakter (-)

$$b_{ij} = \frac{\min_i x_{ij}}{x_{ij}} \cdot 100 \quad (2)$$

pre pevné j ($j \in \{1, \dots, 10\}$), $i \in \{1, \dots, 22\}$). Anomálne (záporné) počty bodov boli prípadne zmenené na nula bodov. Pre každý štát sa určil integrovaný ukazovateľ ako vážený súčet bodov získaný za jednotlivé ukazovatele. Štát s najvyššou hodnotou váženého súčtu bodov získal výsledné poradie 1 atď. až štát s najnižšou hodnotou váženého súčtu bodov získal do výsledného hodnotenia poradie 22.

V rámci metódy normovanej premennej sa hodnoty každého ukazovateľa normovali, pričom sa prihliadalo na charakter ukazovateľa vyjadrený signom (+), resp. (-), resp. normované hodnoty ukazovateľa s charakterom (-) sa navyše „preznamenkovali“. Pre každý štát bol stanovený ako integrovaný ukazovateľ vážený súčet normovaných hodnôt. Štátu

s najvyššou hodnotou váženého súčtu normovaných hodnôt bolo stanovené výsledné poradie 1 atď. až štát s najnižšou hodnotou váženého súčtu normovaných hodnôt bolo priradené poradie 22.

Pri metóde vzdialenosti od fiktívneho objektu sa pracovalo s normovanými hodnotami ukazovateľov, ale do súboru štátov bol zavedený ešte fiktívny štát s hodnotami získanými ako normované hodnoty najlepších zistených hodnôt pre jednotlivé ukazovatele. Pre každý štát bola spočítaná vážená euklidovská vzdialenosť jeho normovaných hodnôt od normovaných hodnôt fiktívneho štátu. Čím bližšie bol daný štát k fiktívnemu objektu, tým lepšie výsledné poradie dosiahol. Štát s najmenšou váženou euklidovskou vzdialenosťou od fiktívneho objektu sa umiestnil na 1. mieste atď. až štát najvzdialenejší od fiktívneho objektu sa umiestnil na 22. mieste.

3. Výsledky

Výsledné poradie sú prezentované v tabuľke č. 3. Vizualne v uvedených poradiach sa nejavia byť veľké rozdiely a vizuálny dojem je doložený skúmaním korelačných vzťahov

Tabuľka 3. Výsledky celkového hodnotenia vybraných štátov Európskej únie

Poradie	Metóda poradi		Bodovacia metóda		Metóda normovanej premennej		Metóda vzdialenosti od fiktívneho objektu	
	Saaty	SPAN	Saaty	SPAN	Saaty	SPAN	Saaty	SPAN
1	Slovensko	Slovensko	Slovensko	Slovensko	Slovensko	Slovensko	Slovensko	Slovensko
2	Rumunsko	Veľká Británia	Rumunsko	Poľsko	Rumunsko	Poľsko	Španielsko	Poľsko
3	Lotyško	Rakúsko	Litva	Cyprus	Poľsko	Cyprus	Poľsko	Veľká Británia
4	Litva	Poľsko	Poľsko	Lotyško	Litva	Luxembursko	Nemecko	Nemecko
5	Poľsko	Lotyško	Lotyško	Litva	Lotyško	Slovensko	Slovensko	Slovensko
6	Bulharsko	Cyprus	Estónsko	Slovensko	Estónsko	Litva	Česká republika	Španielsko
7	Rakúsko	Litva	Bulharsko	Estónsko	Španielsko	Lotyško	Veľká Británia	Luxembursko
8	Nemecko	Nemecko	Česká republika	Rumunsko	Česká republika	Veľká Británia	Estónsko	Cyprus
9	Estónsko	Estónsko	Cyprus	Veľká Británia	Bulharsko	Rumunsko	Maďarsko	Rakúsko
10	Veľká Británia	Rumunsko	Španielsko	Česká republika	Cyprus	Estónsko	Cyprus	Česká republika
11	Cyprus	Luxembursko	Maďarsko	Luxembursko	Maďarsko	Rakúsko	Portugalsko	Francúzsko
12	Portugalsko	Francúzsko	Slovensko	Rakúsko	Slovensko	Nemecko	Rumunsko	Estónsko
13	Španielsko	Slovensko	Portugalsko	Nemecko	Nemecko	Česká republika	Litva	Maďarsko
14	Fínsko	Česká republika	Nemecko	Španielsko	Portugalsko	Španielsko	Rakúsko	Fínsko
15	Česká republika	Španielsko	Rakúsko	Bulharsko	Luxembursko	Francúzsko	Luxembursko	Litva
16	Luxembursko	Dánsko	Veľká Británia	Maďarsko	Rakúsko	Maďarsko	Fínsko	Rumunsko
17	Maďarsko	Portugalsko	Luxembursko	Francúzsko	Veľká Británia	Bulharsko	Lotyško	Lotyško
18	Slovensko	Fínsko	Fínsko	Portugalsko	Fínsko	Portugalsko	Francúzsko	Portugalsko
19	Švédsko	Bulharsko	Francúzsko	Dánsko	Francúzsko	Fínsko	Bulharsko	Belgicko
20	Francúzsko	Švédsko	Švédsko	Fínsko	Švédsko	Dánsko	Belgicko	Dánsko
21	Dánsko	Maďarsko	Dánsko	Švédsko	Dánsko	Švédsko	Švédsko	Bulharsko
22	Belgicko	Belgicko	Belgicko	Belgicko	Belgicko	Belgicko	Dánsko	Švédsko

Zdroj: vlastné spracovanie.

Tabuľka 4. Výsledky testovania (ne)zhody medzi poradiami

Metóda stanovenia váh		Saaty	SPAN	Spolu	Metóda určenia výsledného poradia	Bodovacia	Normovanej premennej	Poradi	Vzdial. od fikt. objektu	Spolu
Saatyho	Priemerný korelačný koeficient	0.689	0.573	0.623	Priemerný korelačný koeficient	0.797	0.843	0.708	0.497	0.699
	Počet korelačných koeficientov	8	8	14	Počet korelačných koeficientov	1	2	2	2	7
	Kendallov koeficient konkordancie	0.741	0.627	0.650	Kendallov koeficient konkordancie	0.922	0.854	0.749	0.742	0.811
	Signifikancia	0.000	0.000	0.000	Signifikancia	0.011	0.023	0.067	0.000	0.000
SPAN	Priemerný korelačný koeficient	0.573	0.759	0.653	Priemerný korelačný koeficient	0.843309	0.662338	0.675607	0.554207	0.68694
	Počet korelačných koeficientov	8	6	14	Počet korelačných koeficientov	2	1	2	2	7
	Kendallov koeficient konkordancie	0.627	0.799	0.678	Kendallov koeficient konkordancie	0.922	0.838	0.777	0.732	0.811
	Signifikancia	0.000	0.000	0.000	Signifikancia	0.011	0.027	0.050	0.000	0.000
Poradi	Priemerný korelačný koeficient	0.70751	0.675607	0.687182	Priemerný korelačný koeficient	0.70751	0.675607	0.687182	0.444684	0.620392
	Počet korelačných koeficientov	2	2	2	Počet korelačných koeficientov	2	2	1	2	7
	Kendallov koeficient konkordancie	0.854	0.838	0.722	Kendallov koeficient konkordancie	0.854	0.838	0.722	0.675	0.675
	Signifikancia	0.023	0.027	0.085	Signifikancia	0.023	0.027	0.085	0.000	0.000
Vzdialenosti od fiktívneho objektu	Priemerný korelačný koeficient	0.497177	0.554207	0.444664	Priemerný korelačný koeficient	0.497177	0.554207	0.444664	0.827216	0.545616
	Počet korelačných koeficientov	2	2	2	Počet korelačných koeficientov	2	2	2	1	7
	Kendallov koeficient konkordancie	0.749	0.777	0.722	Kendallov koeficient konkordancie	0.749	0.777	0.722	0.611	0.611
	Signifikancia	0.067	0.050	0.085	Signifikancia	0.067	0.050	0.085	0.000	0.000

Zdroj: vlastné spracovanie.

medzi výslednými poradiami: najmenší korelačný koeficient je 0.256 a najvyššia hodnota korelačného koeficienta je 0.986. Otázne je, do akej miery sú uvedené poradie zhodné a či

použitá metóda na stanovenie váh alebo použitá metóda viackriteriálneho hodnotenia ovplyvňuje výsledné hodnotenie. V prvom prípade treba konštatovať, že uvedeným navzájom rôznym poradiam zodpovedá priemerná hodnota korelačného koeficientu 0.638 (spočítaná z 28 rôznych párov poradí) a Kendallov koeficient konkordancie 0.651, ktorý sa javí byť štatisticky významným na všetkých bežných hladinách významnosti (signif. 0). V ostatných prípadoch, ako dokladá tabuľka č. 4, tiež nie je pochybovať o zhode medzi poradiami ani v dôsledku použitého iného spôsobu stanovovania váh či stanovenia samotného poradia.

4. Diskusia a záver

Na základe porovnania poradia jednotlivých štátov získaného prostredníctvom uvedených štyroch metód sme zistili, že najlepšie podmienky pre založenie a fungovanie podnikateľských subjektov boli zhruba v roku 2009 v nových členských štátoch, pričom najlepšie podmienky mali byť, prekvapujúco, v Slovenskej republike. Najmenej vhodným miestom pre založenie a fungovanie podnikateľskej jednotky sa javilo byť pre rok 2009 najčastejšie v Belgicku. Napriek tomu, že Belgicko vykazovalo jednu z najvyšších produktívít práce a hrubý domáci produkt na osobu, umiestnilo sa na poslednom mieste, a to hlavne z dôvodu vysokých nákladov na prácu, vysokej minimálnej mzde a vďaka štatisticky nízkej vzdelanostnej úrovni. Aj ďalšie dva staršie členské štáty, Dánsko a Švédsko, dosahovali životnou úrovňou, ale podnikanie sa spájalo s vysokými nákladmi na pracovnú silu. Na záver môžeme konštatovať, že najlepšie podmienky pre založenie a fungovanie nadnárodných foriem obchodných spoločností boli v roku 2009 podľa zvolených aspektov a na základe multikriteriálnych metód v Slovenskej republike. Je dôležité podotknúť, že hodnoty viacerých ukazovateľov pre Slovenskú republiku subsumovali a prenášali dozvuky pozitívnych hospodárskych reforiem vlády prvého a druhého funkčného obdobia Mikuláša Dzurindu, a ak by sa dátový súbor aktualizoval, aj v dôsledku prebiehajúcej (alebo skončivšej sa) hospodárskej krízy by sa Slovenská republika nemusela umiestniť na prvom mieste.

Článok nadväzuje na výskumný projekt VEGA č. 1/0275/09 Implementácia konceptu spoločenskej zodpovednosti podnikov na malé a stredné podniky.

Použitá literatúra

1. DVOŘÁČEK, J. et al. 2006. *Medzinárodné podnikanie: spoločné a nadnárodné podniky*. Žilina: Poradca podnikateľa, 2006. 184 s. ISBN 80-88931-59-2.
2. FICZOVÁ, I., SEDLÁČEK, J., ÚRADNÍČEK, V. 2002. *Finančno-ekonomická analýza podniku: praktikum časť 1*. Banská Bystrica: Občianske združenie Financ, 2002. 208 s. ISBN 80-968702-1-1.
3. HEBÁK, P., HUSTOPECKÝ, J. 1990. *Průvodce moderními statistickými metodami*. Praha: SNTL, 1990. 296 s. ISBN 80-03-00534-5.
4. ÚRADNÍČEK, V. 1998. *Alternatívne prístupy k hodnoteniu postavenia firmy na trhu v podmienkach Slovenskej republiky*. In Acta Facultatis Aerarii Publici. Banská Bystrica: Fakulta financií Univerzity Mateja Bela v Banskej Bystrici, 1998. ISBN 80-8055-145-6. S. 93-102.

Adresa autorov

Martin Balajty, Ing.
Univerzita Mateja Bela v Banskej Bystrici
Ekonomická fakulta
KEMP
Tajovského 10
975 90 Banská Bystrica
martin.balajty@umb.sk

Martin Boďa, Ing. et Bc.
Univerzita Mateja Bela v Banskej Bystrici
Ekonomická fakulta
KKMIS
Tajovského 10
975 90 Banská Bystrica
martin.boda@umb.sk

Vztah spotřeby a objemu dlouhodobých úvěrů domácnostem v ČR The relationship of consumption and the volume of long-term loans to households in the Czech Republic

Lenka Bízová

Abstract: Czech Republic, like other countries in the European Union deals with its debt. The reality is that money is nowadays made up of the banking system and issued into circulation through loans. The article defines the concepts of consumption and total loans to households. Further attention is paid to the techniques of analysis of the economic cycle - the Hodrick-Prescott filter and Bry-Boshan algorithm which are used to find the growth cycle of the total consumption of the Czech Republic and the cyclical development of loans to households. It assessed the relationship of consumption and the volume of loans provided to households and the resulting recommendations for policy-makers.

Key words: economic cycle, consumption, loans to households, correlation.

Klíčová slova: hospodářský cyklus, spotřeba, úvěry domácnostem, korelace.

JEL classification: C5, D12, E32, G19.

1. Úvod

Souvislost mezi spotřebou České republiky a objemem úvěrů v ekonomice je velmi diskutovaným tématem. Již ekonomové rakouské ekonomické školy dávali do souvislosti bankovní úvěry a hospodářský cyklus. V jejich teoriích nalezneme názory, že expanze bankovních úvěrů je příčinou hospodářských cyklů, tedy že cyklické výkyvy ekonomiky jsou způsobeny penězi, hlavně bankovními úvěry [12]. Mezi soudobými zdroji lze zmínit studii Lown a Morgana [11] zabývající se souvislostmi mezi úvěrovým a hospodářským cyklem a taktéž například studii Silvia a Whalla [13], která řeší vliv nebankovních kapitálových půjček na změny v hospodářském cyklu. Zajímavým makroekonomickým ukazatelem, na němž lze modelovat hospodářský cyklus, je celková spotřeba, která je dána do souvislosti s mikroekonomickým ukazatelem dlouhodobých úvěrů domácnostem. Dlouhodobé úvěry domácnostem mají nejvyšší vliv na vývoj celkových úvěrů domácnostem, v 1. čtvrtletí roku 2010 se na celkových úvěrech domácnostem podílely z 96,56 % [2]. Cílem příspěvku je identifikovat charakter vztahu těchto dvou ukazatelů, na jehož základě lze odvodit jejich vzájemné chování v rámci ekonomického vývoje a důsledky pro hospodářskou politiku státu.

2. Metodika

Teorie hospodářského cyklu hovoří a dvou základních přístupech k hospodářskému cyklu, a to o klasickém a růstovém pojetí. Jak uvádí například Burns a Mitchell [4] v klasickém pojetí jsou hospodářské cykly chápány jako opakující se fluktuace v celkové ekonomické aktivitě národů, tedy jde o kolísání úrovně ekonomické aktivity. Růstový cyklus je dle Lucase [10] chápán jako opakující se fluktuace časové řady makroekonomické proměnné okolo svého trendu. Růstový cyklus je založen na dekompozici časových řad zvolené makroekonomické proměnné na trendovou, cyklickou popř. sezónní a nepravidelnou složku. Úvodním krokem analýzy růstového pojetí hospodářského cyklu je aplikace filtrační a následně datovací techniky s cílem identifikovat růstový hospodářský cyklus. Cyklem se pro empirickou část příspěvku rozumí období mezi fázemi vrchol – vrchol [1].

Časovou řadu se rozumí posloupnost věcně a prostorově srovnatelných pozorování, která jsou jednoznačně uspořádána z hlediska času ve směru minulost – přítomnost. Analýzou časových řad se pak rozumí soubor metod sloužících k popisu těchto řad (Hindls a kol., 2006). Takovou časovou řadu lze zapsat ve tvaru:

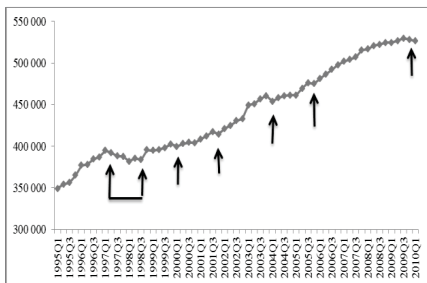
$$y_t = T_t + C_t + \varepsilon_t, t = 1, \dots, n. \quad (1)$$

Cyklická složka C_t vstupní časové řady ukazatele y_t (spotřeba, úvěry domácnostem) je získána jako reziduum po odstranění dlouhodobé složky T_t z původní časové řady pozorování. V tomto příspěvku je pro identifikaci hospodářského cyklu využita filtrační technika Hodrickova-Prescottova filtru [9], datování je provedeno pomocí Bryova-Boshanova algoritmu [3]. Poznamenejme, že při práci s Hodrickovým-Prescottovým filtrem dochází ke zjednodušení vztahu (1) do podoby $y_t = T_t + C_t, t = 1, \dots, n$, kdy je náhodná složka chápána jako součást cyklické složky.

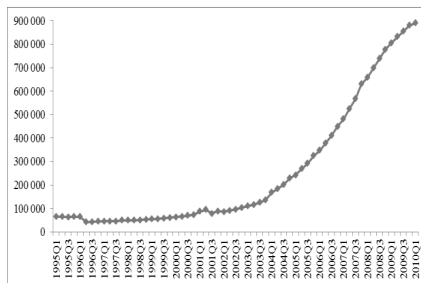
Aplikací výše zmíněných metod jsou získány vstupy pro identifikaci síly závislosti ve vztahu mezi celkovou spotřebou a objemem dlouhodobých úvěrů domácnostem. Míra závislosti je měřena pomocí korelačních koeficientů, které jsou testovány na statistickou významnost [8].

3. Vývoj ukazatelů

Celková spotřeba České republiky je makroekonomickým ukazatelem konečné spotřeby agregátů, včetně rozdělení mezi spotřebou domácností a vlády. Zpracovávané údaje jsou vykazovány v běžných a stálých cenách a zahrnují odpovídající implicitní cenové indexy, jsou sezónně očištěné a ve čtvrtletní frekvenci [6].



Obrázek 1: Vývoj celkové spotřeby ČR



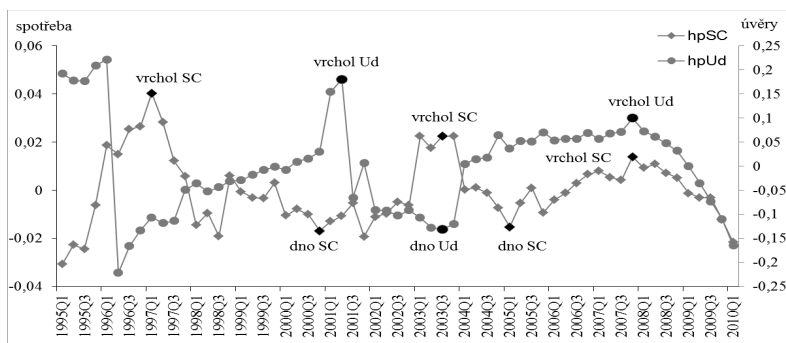
Obrázek 2: Vývoj objemu dlouhodobých úvěrů domácnostem

Na obrázku 1 je znázorněn vývoj celkové spotřeby mezi lety 1995 až 2010. Rostoucí trend značí, jak po celé období dochází k postupnému navyšování objemu spotřeby. Poklesy v rostoucím trendu lze spatřit v období 1997Q1 až 1998Q2, dále dva mírnější ke konci roku 1999 a ve čtvrtletí 2001Q3. Další výraznější pokles rostoucího trendu byl zaznamenán ve čtvrtletí 2003Q4, mírnější v 2005Q3 a poslední od 2009Q3. Veškeré poklesy jsou označeny šipkami. Dlouhodobé úvěry domácnostem (ukazatel mikroekonomický) jsou definovány dle metodického listu České národní banky [5]. Dělení jsou dle několika hledisek. Podle druhového hlediska (typ úvěru), podle měn (CZK), podle příslušnosti klientů-dlužníků do jednotlivých ekonomických subsektorů, podle příslušnosti klientů-dlužníků do jednotlivých odvětví CZ-NACE (do roku 2009 OKEČ) a podle účelu úvěru. Hodnoty úvěrů nejsou kompenzovány o opravné položky, jsou vykazovány v nominální hodnotě, jsou předkládány za banky a pobočky zahraničních bank na území České republiky. Stavby úvěrů jsou za obchodní banky (bez ČNB). Obrázek 2 zachycuje vývoj dlouhodobých úvěrů domácnostem v období let 1995 – 2010. Od roku 2002 lze pozorovat postupně zvyšující se rostoucí

tendenci. Důvodem počátku růstu objemu úvěrů domácnostem v roce 2002 může být označen zejména nově platný zákon 321/2001 Sb., o některých podmínkách sjednávání spotřebitelského úvěru a o změně zákona č. 64/1986 Sb., který nabyl účinnosti 1. 1. 2002 [7]. Zákon ukládá společnostem, které poskytují spotřebitelské úvěry, zakomponovat do smluvních ujednání některé podmínky smlouvy v souladu s právem Evropských společenství. Tento zákon je aktuálně platný.

4. Empirická část

Pro identifikaci růstového cyklu spotřeby a cyklu dlouhodobých úvěrů domácnostem (obrázek č. 3) je v prvním kroku aplikován Hodrickův-Prescottův (HP) filtr. Vzhledem ke čtvrtletní periodicitě dat je zvolena pro HP filtr hodnota parametru $\lambda = 1600$ [3], [9].



Obrázek 3: cykly – body zlomu (vrchol, dno)

Hodnoty růstového cyklu celkové spotřeby (řada hpSC) jsou vyneseny na hlavní vodorovné ose (měřítko vlevo), cyklické hodnoty dlouhodobých úvěrů domácnostem (řada hpUd) jsou vztaheny k vedlejší vodorovné ose (měřítko vpravo). Ve druhém kroku je aplikována datovací technika Bryova-Boshanova algoritmu, jenž je pokročilou metodou generování bodů zlomu [3]. Po aplikaci této techniky identifikujeme body zlomu (vrcholy a dna), které jsou zachyceny taktéž na obrázku 3. U růstového cyklu spotřeby bylo identifikováno celkem pět bodů zlomu – tři vrcholy a dvě dna. První vrchol nastal v 1997Q1, druhý ve čtvrtletí 2003Q3 a třetí 2007Q4. První dno nastalo 2000Q4, druhé v 2005Q1. U cyklického vývoje dlouhodobých úvěrů domácnostem měl Bryaův-Boshanův algoritmus problém identifikovat jakýkoli bod zlomu až do konce roku 2000¹. Celkem byly identifikovány pouze tři body zlomu – dva vrcholy a jedno dno. První fáze vrcholu nastala 2001Q2, druhý vrchol nastal 2007Q4. Fáze dna nastala 2003Q3. Přehledně jsou nalezené body zlomu, zapsány v tabulce 1.

Tabulka 1: Identifikované body zlomu

	Celková spotřeba			Dlouh. úvěry domácnostem			
vrchol	1997Q1		2003Q3		2007Q4	2001Q2	2007Q4
dno		2000Q4		2005Q1			2003Q3

Zdroj : vlastní výpočet

¹ Při použití např. Canovových pravidel byly body zlomu identifikovány již v období 1995 – 2000 (Bízová, 2011)

Tabulka 2: Stanovená časová období cyklů

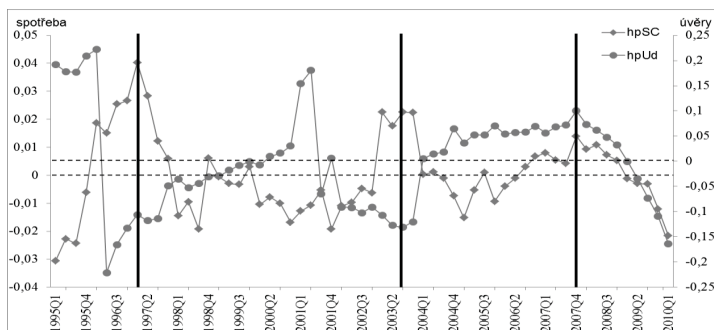
	Celková spotřeba	Dlouhodobé úvěry domácnostem
1. cyklus	1997Q1 - 2003Q3	2001Q2 - 2007Q4
2. cyklus	2003Q4 - 2007Q4	x

Zdroj : vlastní výpočet

Z údajů ve čtvrtletí 2003Q3 vyplývá, že růstový cyklus spotřeby byl aktuálně na svém vrcholu a cyklický vývoj dlouhodobých úvěrů domácnostem se nacházel na svém dnu. Shodná situace nastala u ukazatelů ve čtvrtletí 2007Q4, kdy dosáhly oba svých vrcholů. Tomuto vývoji přispělo sladování ekonomických cyklů před a zejména po vstupu České republiky do Evropské unie. Dle získaných bodů zlomu jsou stanovena časová období jednotlivých cyklů v průběhu sledovaných let 1995 – 2010. Tabulka 2 ukazuje, že byly nalezeny dva plné cykly u růstového cyklu spotřeby a jeden u cyklického vývoje dlouhodobých úvěrů domácnostem.

Na základě výše získaných výsledků byly dále vypočteny korelace filtrovaných hodnot ukazatelů, dle kterých je zjištěna síla závislosti jejich vztahu a statistická významnost. Nejprve je zjištěna síla vztahu za celé sledované období a dále za dílčí cykly. Dílčí cykly jsou brány pouze ty, které jsou stanoveny v tabulce 2 pro celkovou spotřebu.

Vypočtený korelační koeficient $r = -0,3894$ pro celé časové období je statisticky významný (p-hodnota = 0,0019) a značí střední nepřímou závislost mezi pozorovanými veličinami. Pokud docházelo ve sledovaném období k poklesu v objemu dlouhodobých úvěrů domácnostem, pak ve stejném období docházelo k nárůstu hodnot celkové spotřeby, a naopak. Vztah nepřímé závislosti celkové spotřeby a dlouhodobých úvěrů domácnostem určený výpočtem doplňuje obrázek 4. I zde lze pozorovat, že při nárůstu objemu dlouhodobých úvěrů domácnostem, dochází k poklesu celkové spotřeby.

**Obrázek 4: Nepřímá závislost vztahu celkové spotřeby a dlouh. úvěrů domácnostem**

Svislé tučné čáry na obrázku 4 značí konce a začátky stanovených časových období dílčích růstových cyklů celkové spotřeby. Dle těchto dílčích období jsou vypočteny korelační koeficienty (tabulka 4) včetně statistických významností (p-hodnot). Prostřednictvím těchto korelačních koeficientů lze určit sílu závislosti mezi růstovým cyklem celkové spotřeby a cyklickým vývojem dlouhodobých úvěrů domácnostem po částech v časových obdobích, kdy první začíná prvním sledovaným čtvrtletím 1995Q1 a čtvrté končí posledním sledovaným čtvrtletím 2010Q1.

Tabulka 3: Míra závislosti růstového cyklu celkové spotřeby a cyklického vývoje dlouhodobých úvěrů domácnostem dle časových období cyklů celkové spotřeby

Časové období	Dílčí korelační koeficient	Statistická významnost	p-hodnota
1995Q1 – 1996Q4	-0,6976	***	0,0544
1997Q1 – 2003Q3	-0,5607	***	0,0023
2003Q4 – 2007Q4	-0,3972	*	0,1144
2008Q1 – 2010Q1	0,9782	***	0,0000

Z tabulky 3 vidíme, že v prvních třech dílčích obdobích se vztah celkové spotřeby a dlouhodobých úvěrů domácnostem vyznačoval nepřímou závislostí, kdy intenzita závislosti postupně klesá. V prvním období 1995Q1 – 2006Q4 značí korelační koeficient (-0,6976) vyšší sílu nepřímé závislosti, která je však statistická významná až na 10% hladině významnosti. Je tomu tak z důvodu relativně malého počtu vstupních hodnot ($n=8$). Obrázek 4 taktéž potvrzuje, že v tomto období lze hovořit o vyšší nepřímé závislosti ve vztahu mezi cyklickým vývojem dlouhodobých úvěrů domácnostem a růstovým cyklem celkové spotřeby. Ve druhém období 1997Q1 – 2003Q3 značí korelační koeficient (-0,5607) také vyšší sílu nepřímé závislosti, která byla statisticky významná na hladině významnosti 5 %. Při pohledu na toto období (obrázek 4) je zřejmé, že při současném zvýšení (snížení) objemu dlouhodobých úvěrů domácnostem, dochází ke snížení (zvýšení) celkové spotřeby. Pro třetí období 2003Q4 – 2007Q4 byl vypočten korelační koeficient (-0,3972) značící střední sílu nepřímé závislosti, která ale nebyla p-hodnotou (0,1144) potvrzena jako statisticky významná ani na 10% hladině významnosti. V tomto období se dá hovořit o vlivu vstupu České republiky do Evropské unie, v jehož rámci docházelo ke sladování ekonomických cyklů a implementaci evropských zákonných norem, a který je důvodem poklesu nepřímé závislosti mezi celkovou spotřebou a dlouhodobými úvěry domácnostem a přechodu do závislosti přímé. Při prvním pohledu na obrázek 4 lze nabýt dojmu, že se v tomto období jedná o silnější nepřímou závislost – pokud dojde ke snížení (zvýšení) objemu dlouhodobých úvěrů domácnostem, dojde ke zvýšení (snížení) objemu celkové spotřeby. Při důkladnějším pohledu lze zjistit, že od 1. čtvrtletí roku 2005, tedy větší část tohoto období, se pohybují hodnoty obou ukazatelů stejným směrem – pokud došlo ke zvýšení objemu dlouhodobých úvěrů domácnostem, došlo také ke zvýšení celkové spotřeby a naopak. Odlíšnost nastala např. ve čtvrtletí 2005Q4, kdy se ukazatele vyvinuly protichůdně. Ve čtvrtém období 2008Q1 – 2010Q1 vypočtený korelační koeficient (0,9782) značí silnou přímou závislost, která byla potvrzena taktéž p-hodnotou (0,0000) na hladině významnosti 1 %. V tomto období pokračuje sladování ekonomických cyklů. Taktéž se jedná o období, ve kterém došlo k velké hospodářské a finanční krizi. Pokles objemu úvěrů domácnostem v ekonomice koresponduje s poklesem objemu celkové spotřeby (obrázek 4).

5. Závěr

Cílem příspěvku bylo identifikovat charakter vztahu celkové spotřeby a dlouhodobých úvěrů domácnostem a odvodit jejich vzájemné chování v rámci ekonomického vývoje. Na základě provedené analýzy lze říci, že vztah celkové spotřeby a dlouhodobých úvěrů domácnostem se vyznačuje do konce roku 2004 nepřímou lineární závislostí. Tedy pokud se zvýší (sníží) objem dlouhodobých úvěrů v ekonomice, sníží (zvýší) se celková spotřeba České republiky. S přihlédnutím ke grafickému vyjádření lze tvrdit, že dochází ke zpoždění jedné proměnné oproti druhé. Při úvaze, která koresponduje s existujícími ekonomickými teoriemi, lze tvrdit, že jde o zpoždění celkové spotřeby za dlouhodobými úvěry domácnostem v řádu čtyř až devíti čtvrtletí. Po roce 2004 lze hovořit o vlivu sladování ekonomických cyklů a následném paralelním vývoji obou ukazatelů. Vztah celkové spotřeby a dlouhodobých úvěrů domácnostem přechází do přímé závislosti, která se poté od fázi shodných vrcholů ve 4.

čtvrtletí roku 2007 vyznačuje postupným poklesem v objemu obou ukazatelů současně vlivem hospodářské a finanční krize.

Předkládaný příspěvek vznikl za podpory interního grantového projektu 11/2011 “Globální a regionální šoky v procesu evropské integrace”.

6. Literatura

- [1] ARTIS, M., MARCELLINO, M., PROIETTI, T. (2004): Characterising the Business Cycles for Accession countries. CEPR-EABCN Conference of Business Cycle and Acceding Countries, Vienna.
- [2] BÍZOVÁ, L. 2011. Vztah spotřeby a objemu úvěrů domácnostem v ČR. [netištěná diplomová práce]. Brno. Provozně ekonomická fakulta Mendelovi University v Brně. Zápis DP.
- [3] BONENKAMP, J., JACOBS, J., KUPER, G.H. 2001. Measuring Business Cycles in the Netherlands, 1815 – 1913: A comparison of Business Cycle Dating Methods. SOM Research Report, No. 01C25, Systems, Organisation and Management, Groningen. University of Groningen [online].
- [4] BURNS, A.F., MITCHELL, W.C. 1946. Measuring Business Cycles. New York, National Bureau of Economic Research, pp.590, ISBN: 0-870-14085-3;
- [5] ČESKÁ NÁRODNÍ BANKA. 2010. Metodický list. In: www.cnb.cz [online]. [cit. 2011-03-21]. Dostupné z WWW: <http://www.cnb.cz/docs/arady/met_list/tuvob_cs.pdf>
- [6] EUROSTAT. 2010. Quarterly national accounts. In www.ec.europa.eu/eurostat [online]. [cit. 2010-03-21]. Dostupné z WWW: <http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/EN/namq_esms.htm>
- [7] HANZELKA, M. 2002. Komentář k zákonu 321/2001 Sb. In www.finance.cz [online]. [cit. 2011-24-01]. Dostupný z WWW: <<http://www.finance.cz/zpravy/finance/20303-komentar-k-zakonu-321-2001-sb/>>
- [8] HINDLS, R. A KOL. 2006. Statistika pro ekonomy. 7. Vydání. Praha: Profesional Publishing. 415 s. ISBN 80-86946-16-9.
- [9] HODRICK, R. J., PRESCOTT, E.C. 1980. Post-war U.S. Business Cycles: An Empirical Investigation, mimeo, Carnegie-Mellon University, Pittsburgh, PA. 1980, 24pp
- [10] LUCAS, R. E. 1977. Understanding Business Cycles. In: BRUNNER, K., MELTZER, A.H. (eds.): Stabilisation Domestic and International Economy. Carnegie – Orchester Conference Series on Public Policy, Vol. 5, pp. 7–29.
- [11] LOWN, C., MORGAN, D. P. 2004. The Credit Cycle and the Business Cycle: New Findings Using the Loan Officer Opinion Survey. In www.econpapers.repec.org [online]. [cit. 2011-04-16]. Dostupný z WWW: <<http://www.sifr.org/PDFs/sifr-wp27.pdf>>
- [12] MISES, L.V. 2006. Lidské jednání: pojednání o ekonomii. Úprava: Josef Šíma. Praha. Liberální institut. 959 s. ISBN 80-86389-45-6
- [13] SILVIA, J. E., WHALL, B. 2004. Home equity loans and the business cycle. In www.findarticles.com [online]. [cit. 2011-04-16]. Dostupný z WWW: <http://findarticles.com/p/articles/mi_m1094/is_n1_v24/ai_7018978/?tag=content;coll>

Adresa autora:

Lenka Bízová, Bc.

Ústav financí, PEF MU v Brně

Zemědělská I, 613 00 Brno, Česká republika

xbizova0@node.mendelu.cz

**Small-sample inference in the general Gaussian linear regression model
under the assumption of multiplicative heteroskedasticity:
continuation of a Monte Carlo study**

**Inferencia v zovšeobecnenom gaussovskom lineárnom regresnom modeli
v malých výberoch za predpokladu multiplikatívnej heteroskedasticity:
pokračovanie štúdie Monte Carlo**

Martin Boďa

Abstract: The article is a continuation to Monte Carlo studies on the applicability and performance of three methods for the construction of a small-sample confidence region of a linear combination of the regression parameters in the general Gaussian linear regression model with the assumed multiplicative form of heteroskedasticity. A simulation study on the performance of these methods in small samples under several designs is submitted, when multiplicative heteroskedasticity is assumed whilst the variance of the observations of the dependent variable is functionally related to their expectation.

Key words: Kenward-Roger confidence region, conventional confidence region, ML estimation, REML estimation, simulations.

Abstrakt: Článok je pokračovaním štúdií Monte Carlo o využiteľnosti troch metód konštrukcie konfidenčnej oblasti v malých výberoch pre lineárnu kombináciu regresných parametrov v zovšeobecnenom gaussovskom lineárnom regresnom modeli za predpokladu multiplikatívnej heteroskedasticity. Predkladá sa simulačná štúdia o pokrytí týchto metód v malých výberoch pri rôznych simulačných dizajnoch. Predpokladá sa multiplikatívna heteroskedasticita, ale v skutočnosti sú disperzie pozorovaní závisle premennej funkčne závislé na ich strednej hodnote.

Kľúčové slová: Kenwardova-Rogerova konfidenčná oblasť, konvenčná konfidenčná oblasť, odhadovanie maximálnou vierohodnosťou, odhadovanie reziduálnou maximálnou vierohodnosťou, simulácie.

JEL classification: C13, C20.

1. The introduction

The considerations of the article are given to the general Gaussian linear regression model $(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, in which \mathbf{Y} is a Gaussian $(n \times 1)$ random vector with the expectation $\mathbf{X}\boldsymbol{\beta}$ and the covariance matrix $\boldsymbol{\Sigma}$. The design matrix $\mathbf{X} = (\mathbf{x}_1 | \dots | \mathbf{x}_k)'$ is a non-random known $(n \times k)$ matrix of full column rank, and $\boldsymbol{\beta}$ is a $(k \times 1)$ vector of unknown regression parameters. The $(n \times n)$ covariance matrix $\boldsymbol{\Sigma}$ is assumed diagonal with multiplicatively heteroskedastic variance components in the form $\boldsymbol{\Sigma}(\boldsymbol{\alpha}) = \text{diag}\{\exp(\mathbf{z}'_1\boldsymbol{\alpha}), \dots, \exp(\mathbf{z}'_r\boldsymbol{\alpha})\}$, wherein $\mathbf{Z} = (\mathbf{z}_1 | \dots | \mathbf{z}_r)'$ is a non-random known $(n \times r)$ matrix of observations of variables identified to be a cause for heteroskedasticity (with the first column consisting of ones only) and $\boldsymbol{\alpha}$ is an $(r \times 1)$ vector of unknown coefficients. In this setting, the parameters of the model are $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ and both must be estimated; though only the regression parameters $\boldsymbol{\beta}$ are of interest, and it is needful to make inferences of them.

In the estimation of $\boldsymbol{\beta}$ the two-step Aitken (or feasible generalized least squares estimation) procedure is usually employed (cf. e. g. Judge, 1985, p. 422, Judge 1998, p. 352), under which the parameters $\boldsymbol{\alpha}$ are first estimated and the estimate of $\boldsymbol{\beta}$ is constructed as if the covariance matrix were known. For the purpose of inferences about $\boldsymbol{\beta}$ (and its linear combinations as of $\mathbf{L}'\boldsymbol{\beta}$ with \mathbf{L}' being a reasonable $(l \times k)$ fixed matrix), conventionally, a confidence region in the form a Wald-type pivot is constructed, based upon an approximate F distribution with unadjusted degrees of freedom. The employment of an approximate F distribution is justified asymptotically, however, this self-assurance cannot be guaranteed in small samples. Kenward and Roger (1997, 2009) proposed for small-sample inference a scaled Wald-type pivot statistics with an approximate F distribution with adjusted degrees of freedom.

Boďa (2010, 2011) has lately performed Monte Carlo simulations so as to investigate the performance of the conventional confidence region with the confidence regions proposed by Kenward

and Roger constructed about the two-step Aitken estimate of $\boldsymbol{\beta}$ obtained by the maximum likelihood (ML) method or by the restricted maximum likelihood (REML) method. In these simulations, in the model $(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$ the covariance matrix $\boldsymbol{\Sigma}$ was assumed to be governed by the multiplicative heteroskedasticity whilst it might be governed either by (1.) the additive heteroskedasticity, (2.) the mixed heteroskedasticity, (3.) the multiplicative heteroskedasticity, or (4.) the “indefinite” heteroskedasticity, when all observations on the dependent variable are drawn from populations with the same expectation but different variances and heteroskedasticity cannot be parameterized. These simulations, in the setting allowing for a misspecification as to the true form of heteroskedasticity, gave an insight into the robustness of the conventional confidence region and the Kenward-Roger confidence regions.

There, however, is a different and very frequent form of heteroskedasticity which has proven very useful in the theory and practice of econometrics. It is generally assumed that the i -th component σ_i^2 (for $\forall i \in \{1, \dots, n\}$) of the covariance matrix $\boldsymbol{\Sigma}$ can be parametrized as $\sigma_i^2 = h(\mathbf{x}'_i \boldsymbol{\beta}, \boldsymbol{\alpha})$, where h is a reasonable positive function in which the expectation of the i -th observation on the dependent variable Y_i (for $\forall i \in \{1, \dots, n\}$) and an unknown fixed $(r \times 1)$ vector $\boldsymbol{\alpha}$ enters in a non-trivial fashion. Many specification have been proposed for this form of heteroskedasticity, of which most common are – the Box-Hill specifications $\sigma_i^2 = \alpha_1 (1 + |\mathbf{x}'_i \boldsymbol{\beta}|)^{\alpha_2}$, and $\sigma_i^2 = \alpha_1 |\mathbf{x}'_i \boldsymbol{\beta}|^{\alpha_2}$ for $\forall i \in \{1, \dots, n\}$, and – the Jobson-Fuller specification $\sigma_i^2 = |\alpha_1 + \alpha_2 (\mathbf{x}'_i \boldsymbol{\beta})^2|$ for $\forall i \in \{1, \dots, n\}$.

(See a. o. Box-Hill, 1974; Jobson-Fuller, 1980; Carroll a Ruppert, 1982a, 1982b; and Carroll, 1982.) In this article the results of a Monte Carlo simulation study similar to those of Boďa (2010, 2011) are presented aiming to enquire into the performance of the conventional confidence region and the Kenward-Roger confidence regions footed on the ML estimation and on the REML estimation. The true heteroskedasticity is governed by the Box-Hill or Jobson-Fuller specification, all though the incorrect assumption is made that the heteroskedasticity is multiplicative.

2. The simulation study

The simulation study focuses on the performance of the Kenward-Roger methods in situations when the variable causing heteroskedasticity is misidentified as multiplicative and the true heteroskedasticity is that of the Box-Hill specification or the Jobson-Fuller specification.

As in Boďa (2010, 2011), five general simulation designs with a simple linear regression model are explored

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{for } i \in \{1, \dots, n\}, \quad (1)$$

in which Y_i is the explained variable, x_i is the explanatory variable, β_0 and β_1 are the parameters to be estimated, and the disturbance ε_i is distributed as $N(0, \sigma_i^2)$. The dispersion of the disturbance is identified to be dependent on the known variable causing heteroskedasticity, i. e. $\sigma_i^2 = \sigma_i^2(z_i)$, which may be either (most frequently) the explanatory variable x_i itself or (seldom) a different variable $z_i \neq x_i$

For each design the true form of heteroskedasticity appears in the functional form of $(1, z_i)'$ and a nuisance vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$. Two specifications of Box and Hill and one of Jobson and Fuller is considered. The form of heteroskedasticity, designated as the Box-Hill 1 (BH1) specification, runs as

$$\sigma_i^2 = \alpha_1 |\mathbf{x}'_i \boldsymbol{\beta}|^{\alpha_2} \quad \text{for } i \in \{1, \dots, n\}, \quad (2)$$

(with α_1 positive), the Box-Hill 2 (BH2) specification is given by the formula

$$\sigma_i^2 = \alpha_1 (1 + |\mathbf{x}'_i \boldsymbol{\beta}|)^{\alpha_2} \quad \text{for } i \in \{1, \dots, n\}, \quad (3)$$

(with α_1 restricted to be positive), and the Jobson-Fuller (JF) specification is determined by

$$\sigma_i^2 = |\alpha_1 + \alpha_2 (\mathbf{x}'_i \boldsymbol{\beta})^2| \quad \text{for } i \in \{1, \dots, n\}, \quad (4)$$

(requiring that $\boldsymbol{\alpha} \neq \mathbf{0}$).

2.1 The estimation of β , the estimation of α and the construction of the confidence region for L/β

The process of arriving at the conventional confidence region and the Kenward-Roger confidence regions is described exhaustively in the earlier studies (Bod'a, 2010, 2011), and – for it is preserved here – is omitted in this article. The reader is invited to consult the previous studies for details and description. In simulation **L** was set to an identity matrix.

2.2 The simulation framework

Five designs, as before in Bod'a (2010, 2011) were considered, each in 4 modifications according to the parameters of the sampling distribution and the set values of α . The simulation designs and samples were borrowed from the previous studies and are exact the same.

Design A comprised x_i 's and z_i 's in (1) sampled from a uniform population, such that for sub-design Axx $z_i = x_i$ and for sub-design Axz $z_i \neq x_i$. **Design B** was formed of x_i 's and z_i 's drawn from a Gaussian population, such that for sub-design Bxx $z_i = x_i$ and for sub-design Bxz $z_i \neq x_i$. **Design C** (Cxx) consisted in x_i 's distributed uniformly and z_i 's Gaussian and **design D** (Dxx), reversely, was based upon x_i 's Gaussian and z_i 's uniform. The last **design E** was a drawing from a lognormal distribution, again such that for sub-design Exx $z_i = x_i$ and for sub-design Exz $z_i \neq x_i$. Random values of x_i 's and z_i 's were rounded to two decimals, the values for the explanatory variable y_i were simulated according to (1), given the model of heteroskedasticity, and were not rounded. The computations ran first for 10 observations, to which 5 more were added, and the sample of 15 observations was created. To this sample 5 more observations were appended etc.; the simulations were based on samples counting 10, 15, 20, 30 and 50 observations, which were fixed for each evaluation out of 10 000 simulations.

For each design and its modifications, the parameters of the considered distributions and the vector α were determined so that in some designs and their modifications an occurrence of negative values was permitted (for uniform and Gaussian cases) as well as negative values in α were accounted for. The presence of negative values in economic variables is not hypothetical and is not atypical, frequent examples include net assets value, net income or net interests which take oftentimes negative values. The vector of regression parameters was $\beta = (10, 6)'$ in each simulation design and its modifications.

Room here does not suffice to present the full results of the simulation study, therefore, for an illustration, only the results of selected three cases are shown in Table 1 (for designs Axx1, Bxx1, Exx1), and in Table 2 (for designs Axz1, Bxz1, Bxz2). The details and results of the remaining 26 cases are available from the author on request. Under the term coverage the proportion of cases in which simulated data were covered by the respective confidence region statistic at 0.95 confidence level is understood.

Upon interpreting the results, one must beware that assuming multiplicative heteroskedasticity in the case when the variance of the dependent variable is functionally related to its expectation is a serious misspecification and its implications may be expected severe on the performance of any method employed for the set-up of the confidence region. This anticipation turns out true.

In A design (x_i 's uniform – z_i 's uniform) when z_i 's are specified as x_i 's (viz., z_i 's = x_i 's are used) and z_i 's = x_i 's are positive, the confidence regions exhibit at 30-50 observations a good coverage of approx. 0.95. For z_i 's = x_i 's taking also negative values, the confidence regions have a reduced coverage of 0.80 – 0.90 at 30-50 observations, in most cases the REML based estimates outperforming the ML based estimated. When z_i 's are specified different (viz., z_i 's \neq x_i 's are used), the coverage rate is approx. 0.87 – 0.94 at 50 observations, obviously less satisfactory as with z_i 's = x_i 's.

In B design (x_i 's Gaussian – z_i 's Gaussian), in most cases at 30 observations the coverage of the confidence regions is approx. 0.90 – 0.94 and at 50 observations 0.93 – 0.95. In several cases, the coverage rate is clearly lower, which is ordinary when the underlying heteroskedasticity is BH1 and x_i 's and z_i 's are also negative or, generally, when x_i 's and z_i 's take also negative values regardless of the underlying form of heteroskedasticity.

In C design (x_i 's uniform – z_i 's Gaussian), lower coverage rates are displayed in the BH1 specification, and generally higher coverage rates are associated with the JF specification of heteroskedasticity. At 50 observations the coverage rate with the BH1 specification ranges approx.

from 0.80 – 0.90, for the BH2 specification approx. from 0.85 – 0.93, and for the JF specification approx. from 0.90 – 0.93. When both x_i 's and z_i 's are permitted negative values, the coverage is generally lower.

In D design (x_i 's Gaussian – z_i 's uniform), similar as in C design, coverage rates are generally lower for the BH1 specification (approx. from 0.85 – 0.94), and higher for the JF specification (approx. from 0.91 – 0.95), and for possible negative x_i 's and possibly negative z_i 's the coverage rate is sensibly lessened.

Finally, in E design (x_i 's log-normal – z_i 's log-normal) the results vary. For z_i 's = x_i 's, the coverage rates are substantially higher than for z_i 's \neq x_i 's. With z_i 's = x_i 's and 50 observations, the coverage rate runs approx. from 0.90 – 0.95. The ML based confidence regions give a comparatively higher coverage. With z_i 's \neq x_i 's and 50 observations, the coverage rates range approx. between 0.40 – 0.80 for the BH1 specification, approx. between 0.50 – 0.80 for the BH2 specification, and approx. between 0.70 – 0.80 for the JF specification of heteroskedasticity. In contrast, the ML based confidence regions show for z_i 's \neq x_i 's a substantially lower coverage.

In all designs, the Kenward-Roger confidence regions present a better coverage in comparison to the conventional procedure.

Table 1. The results of the selected simulation designs Axx1, Bxx1, Exx1 (Source: the author.)

The coverage results of the selected simulation xx designs												
Method	1997 KR	2009 KR	1997 KR	2009 KR	Standard method		1997 KR	2009 KR	1997 KR	2009 KR		
	REML		ML		REML	ML	REML		ML			
Obs.	Simulation design						TRUE BOX HILL 2 HETEROSKEDASTICITY					
10	Design Axx1						0.8308	0.8949	0.8542	0.8856	0.8711	0.8825
15	$x_i = z_i \sim U(0, 100)$						0.8972	0.9259	0.898	0.9207	0.8975	0.8994
20	for BH1 $\alpha = (1,2)'$, for BH2 $\alpha = (10,0.5)'$, for JF $\alpha = (10,0.50)'$						0.9304	0.945	0.9331	0.9456	0.9293	0.9308
30	$\beta = (10, 6)'$						0.9438	0.953	0.9455	0.9533	0.9413	0.9407
50							0.9505	0.9566	0.952	0.9577	0.9482	0.947
Obs.	TRUE BOX HILL 1 HETEROSKEDASTICITY						TRUE JOBSON-FULLER HETEROSKEDASTICITY					
10	0.9043	0.9371	0.8782	0.9006	0.9053	0.8836	0.844	0.8986	0.8826	0.8925	0.8717	0.8864
15	0.9478	0.9608	0.9422	0.9546	0.9378	0.9274	0.9137	0.9394	0.9151	0.9332	0.9123	0.9123
20	0.9571	0.9669	0.9619	0.9684	0.9506	0.9502	0.9389	0.9529	0.9417	0.9534	0.9363	0.9368
30	0.9647	0.9692	0.97	0.9757	0.9575	0.9609	0.9496	0.9596	0.9517	0.9596	0.9468	0.9468
50	0.9653	0.9686	0.9726	0.9767	0.9598	0.964	0.9591	0.9629	0.9603	0.9653	0.9567	0.957
Obs.	Simulation design						TRUE BOX HILL 2 HETEROSKEDASTICITY					
10	Design Bxx1						0.8045	0.8582	0.8423	0.8736	0.8354	0.8673
15	$x_i = z_i \sim N(2, 2.5)$						0.8828	0.9107	0.8944	0.9124	0.8788	0.8907
20	for BH1 $\alpha = (1,2)'$, for BH2 $\alpha = (10,0.5)'$, for JF $\alpha = (10,0.50)'$						0.9053	0.923	0.909	0.9223	0.8952	0.8994
30	$\beta = (10, 6)'$						0.9266	0.9373	0.931	0.939	0.9141	0.9192
50							0.9408	0.9462	0.9433	0.9487	0.9356	0.9359
Obs.	TRUE BOX HILL 1 HETEROSKEDASTICITY						TRUE JOBSON-FULLER HETEROSKEDASTICITY					
10	0.8065	0.8618	0.8493	0.8787	0.8411	0.874	0.8808	0.908	0.895	0.9141	0.8771	0.8924
15	0.8872	0.9127	0.9017	0.9186	0.8846	0.8998	0.8808	0.908	0.895	0.9141	0.8771	0.8924
20	0.9085	0.9284	0.9168	0.9305	0.8997	0.9078	0.9073	0.9263	0.9114	0.9265	0.8989	0.9031
30	0.9336	0.944	0.9332	0.9419	0.923	0.922	0.9284	0.9378	0.9298	0.9397	0.917	0.9191
50	0.9469	0.9524	0.9474	0.952	0.94	0.9395	0.936	0.9429	0.9383	0.9434	0.9303	0.9307
Obs.	Simulation design						TRUE BOX HILL 2 HETEROSKEDASTICITY					
10	Design Exx1						0.9264	0.9474	0.8525	0.875	0.9247	0.9129
15	$x_i = z_i \sim \Lambda(2, 2.5)$						0.9595	0.9712	0.8696	0.8908	0.9449	0.8986
20	for BH1 $\alpha = (1,2)'$, for BH2 $\alpha = (10,0.5)'$, for JF $\alpha = (10,0.50)'$						0.891	0.9367	0.8741	0.9053	0.8895	0.8726
30	$\beta = (10, 6)'$						0.9245	0.9457	0.9136	0.932	0.9139	0.904
50							0.9467	0.9561	0.945	0.9538	0.938	0.9368
Obs.	TRUE BOX HILL 1 HETEROSKEDASTICITY						TRUE JOBSON-FULLER HETEROSKEDASTICITY					
10	0.8629	0.9047	0.8014	0.8237	0.8635	0.7689	0.9275	0.9527	0.8245	0.8461	0.9273	0.8771
15	0.8538	0.8964	0.8618	0.8808	0.8467	0.7747	0.9491	0.9674	0.8368	0.8623	0.9292	0.8461
20	0.8696	0.9007	0.9266	0.9446	0.8635	0.8711	0.9002	0.9395	0.8842	0.911	0.8904	0.8727
30	0.9057	0.9227	0.9419	0.9556	0.9011	0.9003	0.9299	0.9489	0.9251	0.9401	0.9184	0.9094
50	0.9079	0.9177	0.9459	0.9534	0.9053	0.9086	0.9443	0.9535	0.9464	0.9539	0.9372	0.9355

Table 2. The results of the selected simulation designs $AxzI$, $BxzI$, $ExzI$ (Source: the author.)

The coverage results of the selected xz simulation designs												
Method	1997 KR		2009 KR		Standard method		1997 KR		2009 KR		Standard method	
	REML	ML	REML	ML	REML	ML	REML	ML	REML	ML	REML	ML
Obs.	Simulation design						TRUE BOX HILL 2 HETEROSKEDASTICITY					
10	Design $AxzI$ $x_i \sim U(0, 100)$, $z_i \sim U(0, 100)$, $x_i \neq z_i$ for $BH1 \alpha = (1,2)'$, for $BH2 \alpha = (10,0.5)'$, for $JF \alpha = (10,0.50)'$ $\beta = (10, 6)'$						0.7986	0.8465	0.8674	0.8938	0.789	0.8618
15							0.8687	0.894	0.8894	0.9066	0.85	0.8731
20							0.9104	0.9282	0.9181	0.9322	0.8921	0.9018
30							0.9283	0.9389	0.929	0.9394	0.9174	0.9181
50							0.9369	0.9431	0.942	0.9463	0.9302	0.9311
Obs.	TRUE BOX HILL 1 HETEROSKEDASTICITY						TRUE JOBSON-FULLER HETEROSKEDASTICITY					
10	0.6419	0.7005	0.773	0.8034	0.6277	0.7443	0.765	0.818	0.8437	0.8757	0.7506	0.8336
15	0.7457	0.7807	0.7878	0.8133	0.7219	0.7377	0.8372	0.8663	0.8644	0.8827	0.8179	0.8464
20	0.8374	0.8576	0.8421	0.8576	0.8169	0.7934	0.8861	0.9041	0.8957	0.9105	0.8661	0.8733
30	0.8786	0.8903	0.8792	0.8896	0.8667	0.8291	0.917	0.9258	0.9173	0.9266	0.9081	0.9034
50	0.8712	0.8792	0.8839	0.8914	0.8645	0.8205	0.9189	0.927	0.9223	0.9284	0.9115	0.908
Obs.	Simulation design						TRUE BOX HILL 2 HETEROSKEDASTICITY					
10	Design $BxzI$ $x_i \sim N(2, 2.5)$, $z_i \sim N(2, 2.5)$, $x_i \neq z_i$ for $BH1 \alpha = (1,2)'$, for $BH2 \alpha = (10,0.5)'$, for $JF \alpha = (10,0.50)'$ $\beta = (10, 6)'$						0.8961	0.9206	0.8809	0.9031	0.8771	0.8826
15							0.8955	0.9163	0.9095	0.925	0.8684	0.892
20							0.9169	0.9322	0.9256	0.9382	0.8915	0.9026
30							0.9335	0.9422	0.9342	0.9423	0.9181	0.9205
50							0.9404	0.9473	0.9414	0.9491	0.9307	0.9321
Obs.	TRUE BOX HILL 1 HETEROSKEDASTICITY						TRUE JOBSON-FULLER HETEROSKEDASTICITY					
10	0.8836	0.9111	0.876	0.8951	0.8602	0.8775	0.8922	0.9191	0.884	0.9037	0.8716	0.8831
15	0.9006	0.92	0.9183	0.9316	0.8785	0.8992	0.8969	0.9167	0.917	0.933	0.8721	0.8992
20	0.9168	0.9335	0.926	0.938	0.8964	0.9065	0.9195	0.9338	0.9283	0.9389	0.8982	0.9082
30	0.9323	0.9411	0.9359	0.9443	0.919	0.9213	0.9344	0.9437	0.9361	0.9452	0.9183	0.9208
50	0.9317	0.9376	0.931	0.9368	0.922	0.922	0.9414	0.9484	0.9427	0.9479	0.9339	0.9331
Obs.	Simulation design						TRUE BOX HILL 2 HETEROSKEDASTICITY					
10	Design $ExzI$ $x_i \sim A(2, 2.5)$, $z_i \sim A(2, 2.5)$, $x_i \neq z_i$ for $BH1 \alpha = (1,2)'$, for $BH2 \alpha = (10,0.5)'$, for $JF \alpha = (10,0.50)'$ $\beta = (10, 6)'$						0.8047	0.853	0.7736	0.813	0.8092	0.8087
15							0.3335	0.3688	0.4137	0.4416	0.3277	0.3298
20							0.7739	0.8034	0.8109	0.83	0.7267	0.7683
30							0.8098	0.8304	0.8298	0.8452	0.7675	0.7941
50							0.8579	0.8679	0.865	0.873	0.8502	0.8513
Obs.	TRUE BOX HILL 1 HETEROSKEDASTICITY						TRUE JOBSON-FULLER HETEROSKEDASTICITY					
10	0.3612	0.4114	0.388	0.4249	0.3748	0.38	0.6832	0.7439	0.6626	0.7038	0.6984	0.7041
15	0.3366	0.3744	0.4194	0.4493	0.3296	0.3302	0.629	0.6753	0.6488	0.6812	0.6369	0.6531
20	0.3425	0.3663	0.4791	0.5072	0.3277	0.3039	0.6537	0.686	0.6924	0.7186	0.6066	0.6392
30	0.3793379	0.3975398	0.469847	0.490249	0.3625363	0.3106311	0.6898	0.7128	0.7146	0.7327	0.6472	0.6707
50	0.608	0.6193	0.7886	0.8005	0.5874	0.3939	0.7771	0.788	0.7988	0.8086	0.7645	0.7605

3. The conclusion

The two investigated forms of heteroskedasticity in the Gaussian general linear regression model are possibly most popular in econometric theory and practice. The article considers the situations in which the variance of the dependent variable is functionally dependent on its expectation and yet multiplicative heteroskedasticity is employed and attempts, by Monte Carlo simulations, to evaluate the performance of the two methods for the construction of the confidence region for the regression parameters proposed by Kenward and Roger, and of the conventional method. In the construction of the confidence region both the ML estimation and the REML estimation are studied, and together with the Kenward-Roger methods and the conventional method are juxtaposed. It has been ascertained that

- the two Kenward-Roger confidence regions are comparatively more reliable and exhibit a higher coverage rate when compared with the conventional confidence region,
- the coverage rate of all the three confidence regions is lessened when instead of the independent variable in the design matrix a different variable is identified and employed as the variable causing heteroskedasticity,

- the performance of all the three confidence regions is impaired when either the observations of the independent variable in the design matrix or the observations of the variable identified as causing heteroskedasticity.

Comparing the results of this study with the previous studies (Boďa, 2010, 2011) when the true heteroskedasticity might be additive, mixed, multiplicative, or “indefinite”, it appears that a misspecification of multiplicative heteroskedasticity is – in the case when the variance of the dependent variable is related to the expectation – a more serious issue insomuch as in most situations it impacts the performance of the confidence region more severely.

4. The literature

- [1] BOĎA, M. 2010. Small-sample inference in the general Gaussian linear regression model under the assumption of multiplicative heteroskedasticity based on maximum likelihood estimation In *Forum Statisticum Slovacum*. 2010, vol. 6, iss. 5, pp. 16-26. ISSN 1336-7420.
- [2] BOĎA, M. 2011. Small-sample inference in the general Gaussian linear regression model based on residual maximum likelihood estimation in cases when multiplicative heteroskedasticity is assumed (in press).
- [3] BOX, G. E. P., HILL, W. J. 1974. Correcting inhomogeneity of variance with power transformation weighting. In *Technometrics*, 1974, vol. 16, iss. 3, pp. 385-389. ISSN 0040-1706.
- [4] CARROLL, R. J. 1982. Adapting for heteroskedasticity in linear models. In *The Annals of Statistics*, 1982, vol. 10, iss. 4, pp. 1224-1233. ISSN 0090-5364.
- [5] CARROLL, R. J., RUPPERT, D. 1982a. A comparison between maximum likelihood and generalized least squares in a heteroskedastic linear model. In *Journal of the American Statistical Association*, 1982, vol. 77, iss. 380, pp. 378-882. ISSN 0162-1459.
- [6] CARROLL, R. J., RUPPERT, D. 1982b. Robust estimation in heteroskedastic linear models. In *The Annals of Statistics*, 1982, vol. 10, iss. 2, pp. 429-441. ISSN 0090-5364.
- [7] CARROLL, R. J., RUPPERT, D. 1983. Robust estimators for random coefficient regression models. In SEN, P. K. (ed.) *Contributions to statistics: essays in honor of Norman Lloyd Johnson*. Amsterdam, New York: North-Holland, 1983. 414 pp. ISBN 0 4448 6658 2. S. 81-96.
- [8] JOBSON, J. D., FULLER, W. A. Least squares estimation when the covariance matrix and parameter vector are functionally related. In *Journal of the American Statistical Association*, 1980, vol. 75, iss. 369, pp. 176-181. ISSN 0162-1459.
- [9] JUDGE, G. G. et al. 1985 *The theory and practice of econometrics*. 2nd ed. New York: Wiley, 1985. 1019 p. ISBN 978-0-471-89530-5.
- [10] JUDGE, G. G. et al. 1988. *Introduction to the theory and practice of econometrics*. 2nd ed. New York: Wiley, 1988. 1024 p. ISBN 0-471-62414-4.
- [11] KENWARD, M. G., ROGER, J. H. 1997. Small sample inference for fixed effects from restricted maximum likelihood. In: *Biometrics*, 1997, vol. 53, iss. 3, pp. 983-997. ISSN 0006-341X.
- [12] 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*, 2009, vol. 53, iss. 7, pp. 2583-2595. ISSN 0167-9473.

The author's address

Martin Boďa, Ing. et Bc.

Univerzita Mateja Bela v Banskej Bystrici

Fakulta prírodných vied, Katedra matematiky

Tajovského 40, 974 01 Banská Bystrica

Ekonomická fakulta, Katedra kvantitatívnych metód a informačných systémov

Tajovského 10, 975 90 Banská Bystrica

martin.boda@umb.sk

K validite výsledkov obalovej analýzy dát pri meraní technickej efektívnosti To the validity of DEA results in technical efficiency measurement

Martin Boďa, Viera Roháčová

Abstract: The article points out possible problems concerning the validity of results of data envelopment analysis (DEA) in measuring technical efficiency. The calculations performed upon a simulated data sample with the entirely known multi-product production function indicate that real applications of DEA which are conducted under the absence of the data generating process may not be useful in measuring technical efficiency, and may be misleading.

Key words: data envelopment analysis, multi-product production function, unweighted output-orientated SBM model, technical efficiency score.

Kľúčové slová: obalová analýza dát, viacproduktová produkčná funkcia, nevážený výstupne orientovaný SBM model, skóre technickej efektívnosti.

JEL classification: C65, M11.

1. Úvod

Obalová analýza dát (DEA: *Data Envelopment Analysis*) sa stáva čoraz populárnejšou metódou merania relatívnej efektívnosti produkčných jednotiek, ktorej výhody sa stali hlavným dôvodom jej masívneho rozmachu najmä v posledných rokoch. Napriek tomu, že DEA dáva benchmarkingu istú kvantitatívnu, technicky i ekonomicky podloženú dimenziu, v slovenských podmienkach je problematika merania efektívnosti podnikov a najmä praktická implementácia získaných záverov prostredníctvom DEA len vo svojich začiatkoch.

Z empirických štúdií slovenských autorov sú známe najmä príspevky v oblasti merania efektívnosti poisťovníctva – rozpracovanej napríklad v práci Grmanová (2008); v oblasti merania efektívnosti poľnohospodárskych podnikov – rozpracovanej napríklad v práci Majorová (2007) alebo v oblasti merania efektívnosti pôšt – rozpracovanej napríklad v práci Sedláková (2005). Oblasť merania efektívnosti zdravotníckych systémov prostredníctvom DEA sa venovali napríklad Dlouhý et al. (2007), Novosádová a Dlouhý (2007) či Raguseo a Vlček (2007). V štúdiách autorov Švihlová et al. (2010) a Korholen a Luptáčik (2004) je rozpracovaná problematika uplatnenia DEA pri meraní ekoeфекtívnosti, resp. environmentalistiky. Oblasť uplatnenia DEA pri meraní efektívnosti dopravných systémov sa venovali predovšetkým autori Klieštík (2009) a Klieštík a Birtus (2007).

Hlavná myšlienka DEA spočíva v stanovení miery (spravidla) technickej efektívnosti produkčných jednotiek relatívne vo vzťahu k technickej efektívnosti najlepších produkčných jednotiek v súbore. Každá produkčná jednotka je ohodnocovaná ukazovateľom technickej efektívnosti na intervale $[0,1]$, pričom vyššie hodnoty tohto ukazovateľa indikujú komparatívne vyššiu mieru technickej efektívnosti. Hodnota ukazovateľa efektívnosti 1 poukazuje na to, že príslušná produkčná jednotka v danom súbore produkčných jednotiek patrí k produkčným jednotkám s najlepšou schopnosťou transformovať vstupy na výstupy.

Metodologicky meranie technickej efektívnosti pomocou DEA súvisí s konceptom produkčnej funkcie, ktorou je kvantifikovaná maximálne dosiahnuteľná úroveň výstupov pri daných vstupoch. Produkčná jednotka sa môže nachádzať svojimi vstupmi a výstupmi na produkčnej funkcii (resp. spĺňa jej vzťah), alebo pod úrovňou produkčnej funkcie (a preto neefektívne transformuje svoje vstupy na výstupy). Produkčná funkcia je vlastná dáta generujúcemu procesu a pri praktickej aplikácii je potrebné ju parametricky alebo neparаметricky odhadnúť. V DEA sa *ipso facto* odhaduje produkčná funkcia neparаметricky na základe definície množiny produkčných možností nad napozorovanými vstupmi a výstupmi produkčných jednotiek v dátovom súbore (vzhľadom na charakter produkčného procesu a jeho technologické vlastnosti) a samotný odhad produkčnej funkcie reprezentuje svojimi vstupmi a výstupmi množina produkčných jednotiek, ktoré sú na „obale“ množiny produkčných možností. Produkčné jednotky nachádzajúce sa na „obale“ reprezentujú efektívne produkčné jednotky a je im algoritmom DEA úlohy pridelený ukazovateľ (technickej) efektívnosti 1. Ostatným produkčným jednotkám je pridelený ukazovateľ technickej efektívnosti menší ako 1 podľa ich vzdialenosti k produkčným jednotkám, ktoré sú na „obale“ a sú technicky efektívne.

S týmto je spojená jedna z možných výhod voči DEA ako metóde na meranie (spravidla) technickej efektívnosti produkčných jednotiek. Aproximácia skutočnej produkčnej funkcie je neparаметrická a môže byť veľmi nepresná: čím výraznejšie sa svojimi vstupmi a výstupmi produkčné jednotky nachádzajú pod produkčnou funkciou, tým je skonštruovaný „obal“ dát menšou presnosťou a podhodnoteným estimátorom produkčnej funkcie dáta generujúceho procesu. V dôsledku toho je hodnotenie technickej efektívnosti získané pomocou DEA nadhodnotené, pretože produkčné jednotky sú svojimi vstupmi a výstupmi bližšie odhadnutej „obalovej“ produkčnej funkcie než skutočnej funkcii dáta generujúceho procesu.

Uvedenie si tejto imanentnej vlastnosti DEA motivovalo aj tento článok, v ktorom je postulovaná „výstupne orientovaná“ skutočná produkčná funkcia dáta generujúceho procesu s tromi vstupmi a tromi výstupmi a s charakterom klesajúcich výnosov z rozsahu. Na základe simulovanej dátovej vzorky vstupov a výstupov 30 fiktívnych produkčných jednotiek je pomocou neváženého výstupne orientovaného SBM (*Slack-Based Measure*) modelu zistené hodnotenie technickej efektívnosti a porovnané so skutočnou mierou efektívnosti danej dáta generujúceho procesu. Predpokladá sa, že pri aplikácii DEA na riešenie tejto úlohy bola správne zvolená množina vstupov a výstupov, pričom nie všetky vstupy a výstupy museli byť nevyhnutne použité v rámci DEA, i keď participujú na produkčnom procese. V dôsledku toho boli vstupy a výstupy správne špecifikované, i keď potenciálne všetky použité neboli. Pokiaľ ide o charakter výnosov, hoci dáta generujúci proces sa vyznačoval klesajúcimi výnosmi z rozsahu, bol využitý SBM model s variabilnými výnosmi z rozsahu, čím nedošlo k nesprávnej špecifikácii charakteru výnosov z rozsahu sledovaného produkčného procesu.

V článku je adresovaná pomocou variantného výberu vstupov a výstupov pre implementáciu DEA kľúčna otázka: Poskytuje DEA miery technickej efektívnosti porovnateľné so skutočnou efektívnosťou vlastnou dáta generujúceho procesu?

V nasledujúcich častiach článku je vysvetlený postup generovania dátovej vzorky a spôsob merania efektívnosti pomocou neváženého výstupne orientovaného SBM modelu s variabilnými výnosmi z rozsahu, sú prezentované výsledky a je predložená stručná diskusia k zisteniam.

2. Generovanie dátovej vzorky

Pri generovaní dátovej vzorky sme využili variant viacproduktovej produkčnej funkcie uvažovanej Kleinom (1974, s. 364), ktorú možno v prípade uvažovaných troch výstupov y_1, y_2, y_3 a troch vstupov x_1, x_2, x_3 súčasne používať na ich produkciu zapísať vzhľadom

$$y_1 y_2 y_3^{\alpha} = A x_1^{\beta} x_2^{\delta_1} x_3^{\delta_2}, \quad (1)$$

kde $A, \alpha, \beta, \delta_1, \delta_2, \delta_3$ sú kladné konštanty (porov. aj Theil, 1980, s. 41). Klein (1974, s. 364 a nasl.) využil pri analýze efektívnosti diaľničných komunikácií v Spojených štátoch amerických a odvoláva sa pri definovaní funkčného vzťahu na vlastné empirické výsledky. V predpise (1) boli zvolené hodnoty $A = 1, \alpha = 0.80, \beta = 0.40, \delta_1 = 0.30, \delta_2 = 0.10$ a $\delta_3 = 0.40$, čím sa dosiahla produkčná funkcia

$$y_1 y_2^{0.80} y_3^{0.40} = x_1^{0.30} x_2^{0.10} x_3^{0.40}, \quad (2)$$

ktorou sa deterministicky riadi maximálne dosiahnuteľná úroveň troch výstupov pri danej úrovni uvažovaných troch vstupov fiktívnych produkčných jednotiek. Ide však o maximálne dosiahnuteľnú hladinu výstupov y_1, y_2, y_3 a predpokladá sa, že fiktívne produkčné jednotky dosiahli vo všeobecnosti nižšiu úroveň výstupov, ktorú možno vyjadriť hodnotami $z_1 = \theta_1 y_1, z_2 = \theta_2 y_2$ a $z_3 = \theta_3 y_3$, kde θ_1, θ_2 a θ_3 sú konštanty z intervalu $[0, 1]$. Ak fiktívna produkčná jednotka dosiahla pre nejaký výstup hodnotu $\theta_i = 1$, znamená to, že produkcia príslušného výstupu na základe daných hodnôt vstupov x_1, x_2, x_3 je technicky efektívna. Hodnoty koeficientu $\theta_i < 1$ indikujú technickú neefektívnosť v produkcii príslušného výstupu. V dikcii produkčnej funkcie (2) boli pozorovania vstupov a výstupov pre 30 fiktívnych produkčných jednotiek generované týmto algoritmickým postupom:

1. Najskôr bola generovaná trojica hodnôt x_1, x_2, x_3 , pričom hodnota x_1 bola realizovaná z rozdelenia $LN(1, 0.50)$, hodnota x_2 z rozdelenia $Ro(10, 100)$ a hodnota x_3 bola vybraná náhodne z rozdelenia $N(15, 10)$. Generované hodnoty boli zaokrúhlené na dve desiatinné miesta.
2. Bola spočítaná hodnota výrazu $x_1^{0.30} x_2^{0.10} x_3^{0.40}$ a generovaná hodnota u_1 z rozdelenia $Ro(0.10, 0.50)$ a nezávisle od toho aj hodnota u_2 z toho istého rozdelenia. Na základe toho boli spočítané hodnoty

y_1, y_2 a y_3 využitím pomocných vzťahov $y_1 = u_1 x_1^{0.30} x_2^{0.10} x_3^{0.40}$, $y_2 = (u_2 x_1^{0.30} x_2^{0.10} x_3^{0.40})^{1.25}$ a napokon aj $y_3 = (x_1^{0.30} x_2^{0.10} x_3^{0.40} / y_1 y_2)^{0.80 \cdot 2.5}$.

3. Pre každý potenciálne dosiahnuteľný výstup y bol simulovaný príslušný ukazovateľ technickej efektívnosti θ , z rozdelenia $Ro(0.66, 1.00)$ a spočítaná skutočne dosiahnutá hodnota výstupu $z = \theta \cdot y$, a zaokrúhlená na dve desiatinné miesta. Pre 5 fiktívnych produkčných jednotiek boli zmenené simulované ukazovatele efektívnosti θ , tak, aby boli vo všetkých troch výstupoch fiktívne jednotky technicky efektívne, a okrem toho boli pre 4 ďalšie fiktívne produkčné jednotky upravené ukazovatele technickej efektívnosti θ , v prípade jedného náhodne zvoleného výstupu na plnú technickú efektívnosť.

Týmto postupom boli dosiahnuté pre každú i -tu fiktívnu produkčnú jednotku ($i \in \{1, \dots, 30\}$) pozorovania vstupov $x_{1[i]}, x_{2[i]}$, a $x_{3[i]}$, pozorovania potenciálnych výstupov $y_{1[i]}, y_{2[i]}$, a $y_{3[i]}$, ako aj pozorovania skutočne realizovaných výstupov $z_{1[i]}, z_{2[i]}$, a $z_{3[i]}$. Vzhľadom na to, že je známa produkčná funkcia vrátane parametrov, je možné využiť vzťah (2) na spočítanie skutočnej miery technickej efektívnosti pre ľubovoľnú i -tu fiktívnu produkčnú jednotku ($i \in \{1, \dots, 30\}$) a pre jej dosiahnutú efektívnosť $\theta_{[i]}$ platí

$$\theta_{[i]} = \frac{y_{1[i]} y_{2[i]}^{0.80} y_{3[i]}^{0.40}}{z_{1[i]} z_{2[i]}^{0.80} z_{3[i]}^{0.40}} \quad (3)$$

a ľahko sa dá uvážiť, že nadobúda – vzhľadom na dizajn simulácií, konkrétne bod 3 algoritmu – hodnoty od približne $0.66 \times 0.66^{0.80} \times 0.66^{0.40} \cong 0.40$ až po $1.00 \times 1.00^{0.80} \times 1.00^{0.40} = 1.00$.

Z postupu je zrejmé, že uvažovaný produkčný proces dáta generujúceho procesu je „výstupne orientovaný“ s charakterom klesajúcich výnosov z rozsahu.

3. DEA optimalizačná úloha

Pre exaktnú formuláciu použitého výstupne orientovaného SBM modelu predpokladajme, že hodnoty vstupov a výstupov fiktívnej i -tej produkčnej jednotky ($i \in \{1, \dots, 30\}$) sú reprezentované vektormi $\mathbf{x}_i = (x_{1[i]}, x_{2[i]}, x_{3[i]})'$ a $\mathbf{z}_i = (z_{1[i]}, z_{2[i]}, z_{3[i]})'$ a výstupom zodpovedá vektor sklzov $\mathbf{s}_i = (s_{1[i]}, s_{2[i]}, s_{3[i]})'$. Pre každú fiktívnu produkčnú jednotku sa riešila úloha lineárneho programovania zohľadňujúca všetky možnosti výberu množiny vstupov a výstupov v dikkii skúmaného dáta generujúceho procesu. Postupne bolo skonštruovaných 40 modelov založených na zohľadnení jedného vstupu a dvoch výstupov, dvoch vstupov a jedného vstupu atď. až troch vstupov a troch výstupov (boli vylúčené modely jedného vstupu a jedného výstupu). Optimalizačná úloha založená na všetkých troch vstupoch a troch výstupoch pre produkčnú jednotku i , $i \in \{1, \dots, 30\}$, je určená vzťahom

$$\rho_i(\lambda_i, \mathbf{s}_i) = \left(1 + \frac{1}{3} \sum_{j=1}^{j=30} \frac{s_{j[i]}}{z_{j[i]}} \right)^{-1} =! \min \quad \begin{aligned} \text{pri podmienkach} \quad & \mathbf{s}_i = \sum_{j=1}^{j=30} \{\lambda_j\} \mathbf{z}_j - \mathbf{z}_i \\ & \mathbf{x}_i \geq \sum_{j=1}^{j=30} \{\lambda_j\} \mathbf{x}_j \\ & \mathbf{1}'\lambda = 1, \lambda \geq \mathbf{0}, \mathbf{s}_i \geq \mathbf{0} \end{aligned} \quad (4)$$

Symbol „ \geq “ pri vektore denotuje, že príslušné elementy vektora sú nezáporné a aspoň jeden element je nenulový. Všetky uvažované optimalizačné úlohy obmieňali v (4) iba množinu zvažovaných vstupov a výstupov vo vektoroch \mathbf{x}_i a \mathbf{z}_i a boli pre každú produkčnú jednotku i , $i \in \{1, \dots, 30\}$, založené na napozorovaných hodnotách vstupov a výstupov \mathbf{x}_i a \mathbf{z}_i . Hoci sa uskutočnili v neváženom tvare v snahe kopírovať praktickú aplikáciu DEA pri neznalosti skutočnej produkčnej funkcie (2) dáta generujúceho procesu, váhy by teoreticky bolo možné odvodiť zo znalosti produkčnej funkcie (2) a jej koeficientov. Výsledkom optimalizácie (4) je pre každú i -tu produkčnú jednotku, $i \in \{1, \dots, 30\}$, je hodnota ukazovateľa technickej efektívnosti $\rho_{[i]}$ identifikovaná neváženým výstupne orientovaným SBM modelom. Optimalizácia bola zabezpečená programom DEA Solver Learning version 5. Použitím vzťahu pre skóre technickej efektívnosti v (4) je tiež možné spočítať ukazovateľ technickej efektívnosti, ktorý by mala v ideálnom prípade pre ľubovoľnú i -tu produkčnú jednotku, $i \in \{1, \dots, 30\}$, identifikovať optimalizačná úloha (4) a ktorý je určený vzťahom

$$\rho_i^{me} = \left(1 + \frac{1}{3} \sum_{j=1}^{j=30} \frac{y_{j[i]} - z_{j[i]}}{z_{j[i]}} \right)^{-1}, \quad (5)$$

pričom hodnoty $y_{1[j]} - z_{1[j]}$, $y_{2[j]} - z_{2[j]}$ a $y_{3[j]} - z_{3[j]}$ reprezentujú skutočné hodnoty sklzov, ktoré by mal v ideálnom prípade identifikovať DEA optimalizačná úloha v (4).

4. Metodológia a výsledky

Pre každú z 30 produkčných jednotiek bolo celkom možné stanoviť celkovo 42 skóre technickej efektívnosti: (i.) skutočné skóre efektívnosti θ určené vzťahom (2), (ii.) skutočné skóre efektívnosti prislúchajúce neváženému výstupne orientovanému SBM modelu ρ^{true} určené vzťahom (5) a (iii.) 40 skóre efektívnosti identifikovaných neváženým výstupne orientovaným SBM modelom ako výsledok optimalizačnej úlohy (4). Pri odpovedi na otázky položené v úvode je potrebné porovnávať skóre

Tabuľka 1. Výsledky porovnania skutočného skóre s identifikovaným SBM skóre

Ozn.	# prem.	Model						Porovnanie skutočného skóre s identifikovaným SBM skóre				
		x_1	x_2	x_3	y_1	y_2	y_3	Korelačný koeficient		RMSPE	Wilcoxonov test	
								Pearsonov	Spearmanov		Signif.*	Záver**
1		x			x	x		0.424	0.309	0.559	0.387	zhoda
2		x			x	x	x	0.457	0.376	0.733	0.755	zhoda
3		x				x	x	0.419	0.404	0.755	0.943	zhoda
4			x		x	x		0.582	0.562	0.514	0.202	zhoda
5			x		x	x	x	0.642	0.553	0.543	0.002	ρ^* je vyššie
6			x			x	x	0.454	0.488	0.512	0.003	ρ^* je vyššie
7				x	x	x		0.494	0.472	0.526	0.127	zhoda
8				x	x	x	x	0.604	0.555	0.580	0.044	ρ^* je vyššie
9				x		x	x	0.438	0.516	0.533	0.074	zhoda
10	3	x	x		x			0.534	0.592	0.577	0.003	θ je vyššie
11		x	x			x		0.389	0.242	0.711	0.127	zhoda
12		x	x			x		0.026	0.042	0.873	0.035	ρ^* je vyššie
13		x		x	x			0.360	0.375	0.669	0.020	θ je vyššie
14		x		x		x		0.112	0.061	0.853	0.061	zhoda
15		x		x		x	x	-0.010	-0.007	0.924	0.171	zhoda
16			x	x	x			0.519	0.567	0.615	0.003	θ je vyššie
17			x	x		x		0.416	0.393	0.596	0.245	zhoda
18			x	x		x		0.259	0.218	0.704	0.000	ρ^* je vyššie
19		x			x	x	x	0.506	0.408	0.758	0.030	ρ^* je vyššie
20			x		x	x	x	0.769	0.740	0.371	0.059	zhoda
21				x	x	x	x	0.731	0.759	0.408	0.866	zhoda
22		x	x		x	x		0.529	0.505	0.598	0.010	θ je vyššie
23		x	x		x		x	0.454	0.388	0.735	0.810	zhoda
24		x	x			x	x	0.463	0.410	0.770	0.848	zhoda
25		x		x	x	x		0.324	0.332	0.681	0.026	θ je vyššie
26	4	x		x	x		x	0.458	0.370	0.737	0.848	zhoda
27		x		x		x	x	0.308	0.271	0.902	0.107	zhoda
28			x	x	x		x	0.508	0.556	0.641	0.002	θ je vyššie
29			x	x	x		x	0.542	0.518	0.629	0.195	zhoda
30			x	x		x	x	0.439	0.521	0.532	0.074	zhoda
31		x	x	x	x			0.324	0.360	0.826	0.001	θ je vyššie
32		x	x	x		x		0.242	0.153	0.938	0.002	θ je vyššie
33		x	x	x		x		-0.024	-0.013	0.969	0.206	zhoda
34		x	x		x	x	x	0.499	0.407	0.767	0.026	θ je vyššie
35		x		x	x	x	x	0.455	0.389	0.877	0.003	θ je vyššie
36			x	x	x	x	x	0.671	0.697	0.481	0.361	zhoda
37		x	x	x	x	x		0.297	0.338	0.891	0.000	θ je vyššie
38		x	x	x		x	x	0.396	0.330	0.844	0.665	zhoda
39		x	x	x		x	x	0.384	0.310	0.942	0.034	θ je vyššie
40	6	x	x	x	x	x	x	0.417	0.281	0.922	0.003	θ je vyššie

* Obojstranná p-hodnota. ** Výsledok zodpovedajúci nulovej hypotéze alebo rezultujúcej alternatívnej hypotéze.

Zdroj: vlastné spracovanie.

efektívnosti pre všetkých 30 fiktívnych produkčných jednotkách v rámci každej optimalizačnej úlohy. Výsledky z porovnania skutočného skóre efektívnosti θ a skóre efektívnosti identifikovaného DEA úlohou sú obsahom tabuľky č. 1 a výsledky z porovnania ideálneho skóre efektívnosti zodpovedajúceho DEA úlohe ρ^{true} so skóre efektívnosti identifikovaného DEA úlohou sú prezentované v tabuľke č. 2.

Tabuľka 2. Výsledky porovnania ideálneho SBM skóre s identifikovaným SBM skóre

Ozn.	# prem.	Model						Porovnanie ideálneho SBM skóre s identifikovaným SBM skóre				
		x_1	x_2	x_3	y_1	y_2	y_3	Korelačný koeficient			Wilcoxonov test	
								Pearsonov	Spearmanov	RMSPE	Signif.*	Záver**
1		x			x	x	0.412	0.352	0.341	0.002	ρ^* je vyššie	
2		x			x	x	0.390	0.302	0.531	0.007	ρ^* je vyššie	
3		x				x	0.404	0.390	0.494	0.009	ρ^* je vyššie	
4			x		x	x	0.581	0.549	0.325	0.001	ρ^* je vyššie	
5			x		x	x	0.603	0.511	0.575	0.000	ρ^* je vyššie	
6			x			x	0.453	0.440	0.567	0.000	ρ^* je vyššie	
7				x	x	x	0.514	0.461	0.302	0.003	ρ^* je vyššie	
8				x	x	x	0.569	0.501	0.549	0.000	ρ^* je vyššie	
9				x		x	0.464	0.504	0.530	0.000	ρ^* je vyššie	
10	3	x	x		x		0.499	0.530	0.287	0.079	zhoda	
11		x	x			x	0.360	0.239	0.354	0.037	ρ^* je vyššie	
12		x	x			x	0.006	0.007	0.713	0.000	ρ^* je vyššie	
13		x		x	x		0.360	0.089	0.321	0.086	zhoda	
14		x	x			x	0.115	-0.064	0.390	0.262	zhoda	
15		x		x		x	-0.053	0.335	0.702	0.000	ρ^* je vyššie	
16			x	x	x		0.487	0.378	0.293	0.309	zhoda	
17			x	x		x	0.423	0.194	0.342	0.013	ρ^* je vyššie	
18			x	x		x	0.244	0.491	0.731	0.000	ρ^* je vyššie	
19		x			x	x	0.466	0.374	0.410	0.309	ρ^* je vyššie	
20			x		x	x	0.757	0.688	0.447	0.000	ρ^* je vyššie	
21				x	x	x	0.734	0.716	0.394	0.001	ρ^* je vyššie	
22		x	x		x	x	0.496	0.312	0.298	0.072	zhoda	
23		x	x		x	x	0.388	0.396	0.530	0.008	ρ^* je vyššie	
24		x	x			x	0.438	0.362	0.492	0.017	ρ^* je vyššie	
25		x		x	x	x	0.335	0.301	0.323	0.116	zhoda	
26	4	x		x	x	x	0.401	0.286	0.518	0.014	ρ^* je vyššie	
27		x		x	x	x	0.284	0.386	0.487	0.299	zhoda	
28			x	x	x	x	0.490	0.438	0.301	0.371	zhoda	
29			x	x	x	x	0.480	0.508	0.557	0.001	ρ^* je vyššie	
30			x	x		x	0.464	0.620	0.528	0.000	ρ^* je vyššie	
31		x	x	x	x		0.300	0.128	0.315	0.734	zhoda	
32		x	x	x		x	0.218	-0.068	0.387	0.586	zhoda	
33		x	x	x		x	-0.065	0.254	0.711	0.001	ρ^* je vyššie	
34		x	x		x	x	0.455	0.369	0.411	0.417	zhoda	
35		x		x	x	x	0.440	0.498	0.415	0.813	zhoda	
36			x	x	x	x	0.645	0.311	0.407	0.012	ρ^* je vyššie	
37		x	x	x	x	x	0.259	0.266	0.350	0.417	zhoda	
38		x	x	x	x	x	0.339	0.302	0.534	0.030	ρ^* je vyššie	
39		x	x	x		x	0.344	0.468	0.498	0.750	zhoda	
40	6	x	x	x	x	x	0.384	0.261	0.437	0.629	zhoda	

* Obojsmerná p-hodnota.

** Výsledok zodpovedajúci nulovej hypotéze alebo rezultujúcej alternatívnej hypotéze.

Zdroj: vlastné spracovanie.

Pri porovnávaní bola zhoda skóre efektívnosti posudzovaná pomocou Pearsonovho korelačného koeficientu, Spearmanovho korelačného koeficientu a priemernej percentuálnej kvadratickej chyby (RMSPE: *root mean square percentage error*). V prípade, že pomocou DEA v dikcii sledovaného modelu možno správne zmerať technickú efektívnosť skúmaných fiktívnych produkčných jednotiek, Pearsonov korelačný koeficient by mal vykazovať kladné hodnoty zrejme nie veľmi odlišné od 1. Ak by sme hypotetizovali, že DEA nedokáže kvantifikovať správne technickú efektívnosť skúmaných fiktívnych produkčných jednotiek, je možné, že aspoň v hodnotách skóre efektívnosti zachováva ich relatívne poradie, v tomto prípade by mal byť Spearmanov korelačný koeficient kladný a blízky 1. Podobne nízke hodnoty RMSPE blízke 0 by mali indikovať vyššiu schopnosť DEA merať technickú efektívnosť. Na posúdenie zhody medzi skutočným skóre θ a identifikovaným skóre, resp. ideálnym skóre ρ^{inc} a identifikovaným skóre bol ešte použitý Wilcoxonov párový test, v rámci ktorého sa testovala zhoda posudzovaných skóre technickej efektívnosti (nulová hypotéza) a ktorého alternatívna hypotéza vyplynula z dát.

5. Diskusia a záver

Je zrejme – pri porovnaní so skutočnou mierou θ technickej efektívnosti a ideálnou mierou technickej efektívnosti ρ^{inc} –, že DEA v tomto jednoduchom prípade nebola schopná zmerať

technickú efektívnosť vlastnú realite dáta generujúceho procesu, čo dokazujú pomerne variabilné a pomerne nízke hodnoty Pearsonovho aj Spearmanovho korelačného koeficientu či značne vysoké hodnoty ukazovateľa RMSPE. Skóre technickej efektívnosti získané pomocou DEA sa nielen odlišujú od skutočných hodnôt skóre technickej efektívnosti θ či skóre technickej efektívnosti ρ^{true} ideálne identifikovaného DEA úlohou, ale nedokážu ani dobre identifikovať „poradie“ technickej efektívnosti sledovaných fiktívnych produkčných jednotiek. Napriek tomu Wilcoxonov párový test nedokázal identifikovať v niektorých prípadoch rozdiel, ale z tabuľky č. 2 je zrejme, že v porovnaní so skóre ρ^{true} ideálne identifikovaným DEA úlohou DEA často nadhodnocuje skóre a dáva „optimistickejší“ pohľad na technickú efektívnosť. Tabuľka č. 1 celkom nepotvrďuje tieto tvrdenia, pretože v niektorých prípadoch, keď nedošlo k statistickej zhode skutočného skóre θ a skóre identifikovaného v DEA, v niektorých vzácných prípadoch pri nižšom počte premenných vstupujúcich do DEA, bolo skutočné skóre θ nižšie a skóre technickej efektívnosti identifikované DEA bolo skreslené smerom nadol.

Príspevok je čiastočným výstupom grantu UGA Perspektívy praktickej aplikácie modelov obalovej analýzy dát (DEA) pri meraní a hodnotení efektívnosti podnikov.

Použitá literatúra

1. GRMANOVÁ, E. 2008. Hodnotenie vybraných ukazovateľov poisťovní v ich vzájomnej interakcii. In Biatec, 2008, roč. 16, č. 6. s. 15-18. ISSN 1335-0900.
2. KLIEŠTIK, T. 2009. Kvantifikácia efektivity činnosti dopravných podnikov pomocou data envelopment analysis. In E+M. Ekonomie a management, 2009, č. 1, s. 133-145. ISSN 1212-3609.
3. KLIEŠTIK, T., BIRTUS, M. 2007. Evaluation statically effectiveness activities of road transportation companies by means of data envelopment analysis and dynamic effectiveness by means of Malmquist index. In Vadyba Management, 2007, roč. 2, č. 15, s. 9-13. ISSN 1822-3133.
4. KLEIN, L. R. 1974. A textbook of econometrics. 2. vyd. Englewood Cliffs (New Jersey, USA): Prentice-Hall, 1974. 436 s. ISBN 0-13-912832-8.
5. KORHOLEN, P. J., LUPTÁČIK, M. 2004. Eco-efficiency analysis of power plants: An extension of data envelopment analysis. In European Journal of Operational Research, 2004. č. 154., s. 437-446. ISSN 0377-2217.
6. MAJOROVÁ, M. 2007. Hodnotenie produktivity poľnohospodárskych podnikov na Slovensku v čase- aplikácia Malmquistových indexov. In Mladá veda 2007. Medzinárodná vedecká konferencia mladých vedeckých pracovníkov a doktorandov. Račkova dolina. Nitra: Slovenská poľnohospodárska univerzita v Nitre, 2007, s. 323-328. ISBN 978-80-8069-970-3.
7. NOVOSADOVA, I., DLOUHÝ, M. 2007. Hodnotenie efektívnosti produkčných jednotiek vo verejnom sektore - príklad hodnotenia liečební dlhodobó chorých. In Theoretical and Practical Aspects of Public Finance, 2007, s. 52-60. ISBN 80-245-1032-4.
8. RAGUSEO, D., VLČEK, P. 2007. Health care in Europe- a multicriteria approach. In International Archives, Vol. 70, No. 3, 2007, s. 46-55. ISSN 0323-9004.
9. THEIL, H. 1980. The system-wide approach to microeconomics. Chicago: The University of Chicago Press, 1980. 260 s. ISBN 0-226-79437-7.
10. SEDLÁKOVÁ, J. 2005. Data envelopment analysis and its use in communications. In Transcom, Economics and Management, Section 2, jún 2005, s. 117-120. ISBN 80-8070-414-7.
11. ŠVIHLOVÁ, D., KRAJ, P., HUŽVÁR, M. 2010. Analysis of economic-environmental disparities in Slovakia using data envelopment analysis. In Proceedings of the International Scientific Conference AMSE 2010, Demänovská dolina, August 26-29, 2010. Banská Bystrica, Občianske združenie Financ, 2010. ISBN 978-80-89438-02-0.

Adresa autorov

Martin Bod'a, Ing. et Bc.
Univerzita Mateja Bela v Banskej Bystrici
Ekonomická fakulta / Fakulta prírodných vied
KKMIS / KM
Tajovského 10 / Tajovského 40
975 90 / 974 01 Banská Bystrica
martin.boda@umb.sk

Viera Roháčová, Ing.
Univerzita Mateja Bela v Banskej Bystrici
Ekonomická fakulta
KKMIS
Tajovského 10
975 90 Banská Bystrica
viera.rohacova@umb.sk

Prognóza a verifikácia výkonov v mužských atletických disciplínach do OH v Londýne Prognosis and verification performance in male athletic disciplines to the Olympic Games in London

Jaroslav Broďáni

Abstract: The study verifies forecast the world's best performance in 24 male athletic disciplines established in 2009. On the basis of additional time series on the world's best performances in 2010, provides new projections for the period until the Olympic Games in London, the 2012th.

Key words: track and fields, forecast, verification

Kľúčové slová: atletika, prognóza, verifikácia

JEL classification: L83 - Sports

1. Úvod

Pri výpočtoch prognóz sú spravidla využívané počítače s príslušným programom. Súčasťou výpočtov teoretických kriviek rastu výkonnosti sú informácie o štandardnej chybe odhadu, koeficientu korelácie a determinácie. Na základ týchto informácií sú vybrané najvhodnejšie funkcie. Výsledky výpočtov sú spravidla podporené dotazníkmi od špecialistov v danom športe. Súčasťou analýzy je aj metóda rozhovoru, ktorej výsledky sú využité pri spracovaní prognózy.

Výpočet a zostavovanie prognóz pre budúce roky spravidla predchádza vyhodnocovanie úspešnosti prognóz. Spravidla je použitý vzťah, kedy prognóza predstavuje 100 %, pričom rozdiel medzi prognózou a skutočnými výkonnosťami sa vyjadruje percentuálnou odchýlkou.

Významnú úlohu pri spresňovaní prognóz má ich spätná verifikácia. O spoľahlivosti, pravdivosti, presnosti prognóz môžeme hovoriť až po samotnej realizácii v budúcnosti. Snaha posúdiť spoľahlivosť prognóz už počas spracovania vedie k ich verifikácii. Z verifikačných metód využívame: matematicko-logické metódy, dosiahnutie výsledku prognózy iným postupom, zdvojením dôkazov atď. Verifikácia plní účel poznania v zmysle jeho prehĺbenia aj keď nemôžeme dať jednoznačnú odpoveď na otázku presnosti alebo pravdivosti prognózy (Korček, 1983; Měkota - Cuberek, 2007; Moravec, 2007).

Verifikáciu delíme na priamu a nepriamu (Tilinger, 2004). Priama verifikácia dosahuje rovnakých výsledkov a záverov prognózy ale inou prognostickou metódou. Nepriama verifikácia preveruje predpovede inými cestami, logicky, matematicky a pod.

Pri prognózovaní v športe sa odporúčajú tieto základné spôsoby verifikácie:

- Odkaz na prognózu v rovnakej problematike, spracovanú a doloženú v minulosti,
- Konzekventná verifikácia cestou logických a matematických následkov už známych prognóz.
- Zdvojovanie dôkazov k prognóze, postavenie otázky inou formuláciou,
- Verifikácia inverznou cestou s využitím metód najmenších štvorcov, pričom sa snažíme nájsť regresnú rovnicu pre radu významov prognóz,
- Verifikácia preukázaním nereálnosti (voľba oponentov, ktorí sa zamerajú na výhradné argumenty nereálnosti prognózy)
- Minimalizácia systematických chýb založená na klasifikácii zdrojov chýb.

Verifikácia je spojená s voľbou prognostickej metódy. Čím exaktnejšia je metóda, tým dokonalejšia je verifikácia. Obdobne aj presnosť a početnosť vstupných údajov, ovplyvňujú

presnosť vybranej metódy. Verifikácia je spojená aj s dĺžkou prognózovaného obdobia (Turek - Ružbarský, 2001).

2. Problém

Cieľom štúdie je verifikácia prognóz najlepších svetových výkonov v 24-roch mužských atletických disciplínach stanovených v roku 2009 a stanovenie nových prognóz na obdobie do konania Olympijských hier v Londýne v roku 2012.

3. Metodika

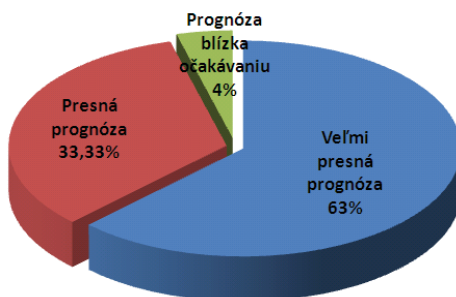
Pri stanovení prognóz na obdobie rokov 2010 - 2012 využívame časový rad údajov najlepších mužských atletických výkonov, ktorý sa pohyboval do roku 2009. Pri vyhladzovaní údajov a stanovení prognózovaných hodnôt sme použili metódy regresie a extrapolácie. Pre odhad a výpočet prognózovaných hodnôt sme použili grafickú a číselnú metódu realizovanú v programe MS Excel. Pri výbere regresnej funkcie bolo taktiež prihliadané na úroveň svetového rekordu, úroveň svetovej výkonnosti v posledných rokoch a v úvahu boli brané názory trénerov špecialistov.

Realizácia verifikácie spočívala v porovnávaní skutočnej hodnoty a percentuálnej odchýlky od stanovenej prognózy, pričom prognóza predstavuje 100 % (Tilinger, 2004). Pri úplnej zhode prognózy so skutočným výkonom považujeme diferenciu, ktorá nepresiahla 1 % za veľmi presnú prognózu. Rozdiel do 2 % považujeme z praktického hľadiska za presnú predpoveď. Skutočné hodnoty, ktoré sa nelíšia od prognóz viac ako o 3 % hodnotíme ako „blízke očakávania“. Väčší rozdiel hodnotíme ako nezgodu. Hodnotenie prognóz je doplnené o matematický rozdiel „d“ a slovné hodnotenie.

Po posúdení prognóz z roku 2009 boli stanovené nové prognózy na obdobie rokov 2011 - 2012, pričom bol časový rad doplnený o výkony v roku 2010. Prognózy sú doplnené o regresnú rovnicu a index spoľahlivosti R^2 .

4. Výsledky

Analýzou najlepších mužských atletických výkonov do roku 2009 a s prihliadnutím na názory vybraných atletických špecialistov v daných disciplínach sme stanovili prognózy na obdobie 3 rokov (2010-2011-2012) v bežeckých, chodeckých, skokanských, vrhačských disciplínach a v mužskom desaťboji (tab. 2). Podrobnosti k analýze trendu vývoja svetových výkonov a stanovenie prognóz boli publikované v štúdií Broďáni (2010).



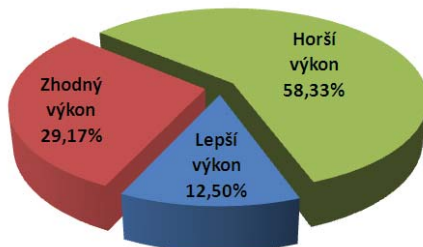
Graf 1: Verifikácia úspešnosti prognóz z roku 2009

S ročným odstupom sme mali možnosť posúdiť úspešnosť našich prognóz s realitou dosiahnutých výkonov v roku 2010. S prihliadnutím na uvádzanú metodiku hodnotenia prognóz sa ukázalo, že 63 % prognóz splnilo kritérium „Veľmi presná prognóza“ s rozdielom

medzi prognózou a skutočným výkonom do 1 %. Medzi 6-timi najpresnejšími prognózami s odchýlkou do 20 % sme zaznamenali disciplíny ako vrh guľou (0,04 %), beh na 200 m (0,05 %), beh na 400 m prekážok (0,08 %), maratón (0,15 %), hod diskom (0,15 %) a štafeta na 4x100m (0,16 %).

Do kritéria „presná prognóza“ s rozdielom medzi prognózou a skutočným výkonom do 2 % spadá 33,33 % prognóz (7 disciplín).

Kritériom do 3 % „Prognóza blízka očakávaniu“ splnilo 4 % predpovedí. Medzi prognózované výkony s rozdielom do 3 % spadala iba 1 mužská atletická disciplína, a to hod kladivom. Prognózy s rozdielom väčším ako 3 % sme nezaznamenali.



Graf 2: Porovnanie nových prognóz s prognózami z roku 2009

Po doplnení údajov do časových radov o najlepšie svetové výkony v roku 2010, sme opätovne prehodnotili predpovede na zvyšné obdobie do OH v Londýne 2012 (tab. 2). V porovnaní s prognózami z roku 2009, sú nové prognózy na obdobie predolympijského a olympijského makrocyclov prevažne skeptickejšie. Iba v troch prípadoch (12,5 %) predpovedáme lepšie výkony ako pri prognózach z roku 2009. Jedná o disciplínu beh na 200 a 800 metrov a v chôdzi na 50 km. Zhodné výkony môžeme očakávať v 29,17 % prípadoch, resp. v 58,33 % prípadoch predpovedáme slabšie výkony.

Z pohľadu svetových rekordov (tab. 1 a 2) v spomínanom predolympijskom a olympijskom období neočakávame prekonanie žiadneho mužského svetového rekordu. Najlepšie svetové výkony sa svojou úrovňou približia iba vo vytrvalostných disciplínach (10000 m, Maratón a chôdzi na 20 km).

Tabuľka 1: Svetové rekordy v mužských atletických disciplínach

Disciplína	Svetový rekord k 31.12.2010	Disciplína	Svetový rekord k 31.12.2010
100 m [s]	9,58	Hod kladivom [m]	86,74
200 m [s]	19,19	Skok do výšky [m]	2,45
110 m prekážok [s]	12,87	Skok do diaľky [m]	8,95
400 m prekážok [s]	46,78	Skok o žrdi [m]	6,14
400 m [s]	43,18	Trojskok [m]	18,29
800 m [s]	01:41,1	Vrh guľou [m]	23,12
1500 m [min]	03:26,0	3 000 m prekážok [min]	07:53,6
5000 m [min]	12:37,4	20 km chôdza [hod]	1:17:16
10000 m [min]	26:17,5	50 km chôdza [hod]	3:34:14
Maratón [hod]	2:03:59	Štafeta 4x100 m [s]	37,1
Hod diskom [m]	74,08	Štafeta 4x400 m [min]	02:54,3
Hod oštepom [m]	98,48	Desaťboj [body]	9026

5. Závěry

Verifikácia prognóz umožnila spresniť úroveň najlepších mužských atletických výkonov na obdobie londýnskeho predolympijského a olympijského makrocyklu.

V sledovanom období neočakávame prekonanie svetových rekordov, maximálne priblíženie najlepších svetových výkonov k svetovým rekordom vo vytrvalostných disciplínach.

Prognózovanie a verifikáciu predpovedí musíme chápať ako nepretržitý proces a poukazuje na vyššiu úspešnosť predpovedí s krátkodobým časovým horizontom.

Úspešnosť krátkodobej prognózy závisí od úzkej spolupráci s odborníkom (prognostikom a znalcom) v analyzovanej oblasti atletickej disciplíny.

Dynamika najlepších svetových atletických výkonov má afinitu k jednotlivým obdobiam olympijského cyklu, intraindividuálnej dynamike výkonnosti najlepších atlétov sveta, kalendára atletických podujatí a ich dôležitosti, kvalitnejšej detekcii zakázaných prostriedkov, kvalite využívaných materiálov a dovolených podporných prostriedkov, atď.

6. Literatúra

- [1] BRODĀNI, J. 2010. Prognóza najlepších svetových výkonov v mužských atletických disciplínach do roku 2012. In *Exercitatio Corpolis Motus Salus* : Slovak journal of sports science. ISSN. 1337-7310. Roč. 2. Č. 1 (2010).
- [2] KORČEK, F. 1983. Prognózovanie v športe. In *Teória športu*. Bratislava : UK, 1983, s.97-112.
- [3] MĚKOTA, K. - CUBEREK, R. 2007. Prognózování sportovní výkonnosti. In *Pohybové dovednosti – činnosti – výkony*. Olomouc: UP, 2007. S. 135-142.
- [4] MORAVEC, R. 2007. Prognózovanie a modelovanie v športe. In *Teória a didaktika výkonnostného a vrcholového športu*. Bratislava: ICM AGENCY, s. 216-222.
- [5] PUPIŠ, M. 2008. Vývoj atletickej chôdze na 20 km na vrcholných podujatiach v rokoch 1964-2008. In *ATLETIKA 2008*. Nitra : UKF, s. 114.
- [6] TILINGER, P. 2004. Prognózování vývoje výkonnosti ve sportu. Praha : Karolinum, 2004, 167 s.
- [7] TILINGER, P. - KOVÁŘ, K. 2000. The developmental and performance trends in Olympic track and field. In *2000 Pre-Olympic Congress. International Congress on Sport Science, Sports Medicine and Physical Education*. Brisbane : Australia, 2000. p. 212.
- [8] TUREK, M. - RUŽBARSKÝ, P. 2001. Športová prognóza v praxi. Prešov: PF PU Prešov, SVSTVŠ, 2001, 83s.

Adresa autora:

Jaroslav BrodĀni, doc. PaedDr. PhD.

KTVŠ PF UKF Nitra

Tr. A. Hlinku 1

949 74 Nitra

jbrodani@ukf.sk

Príspevok je súčasťou grantu MŠ SR VEGA1/0248/11: BrodĀni, J. a kol.: Prognóza svetových výkonov v mužských a ženských atletických disciplínach do OH 2012 v Londýne. Nitra : KTVŠ PF UKF, Riešené 2011-2013.

Tabuľka 2: Verifikácia prognóz mužských atletických výkonov z roku 2009 a stanovenie nových prognóz na obdobie do OH v Londýne

Disciplína	Prognózy výkonov z roku 2009				Posúdenie prognóz				Nové prognózy			Regresné parametre	
	2010	2011	2012	Realita 2010	Rozdiel "d"	Ochýľka %	Slovom	2011	2012	R ²	Rovnica	R ²	
100 m [s]	9,75	9,73	9,71	9,78	0,03	0,31	veľmi presná prognóza	9,74	9,73		$y = -0,0074x + 10,087$	0,8042	
200 m [s]	19,55	19,52	19,48	19,56	0,01	0,05	veľmi presná prognóza	19,38	19,32		$y = -3E-05x^2 + 0,0014x^2 - 0,023x + 20,029$	0,409	
110 m prekážok [s]	12,93	12,93	12,93	12,89	0,04	0,31	veľmi presná prognóza	12,93	12,93		$y = -0,12ln(x) + 13,354$	0,5662	
400 m prekážok [s]	47,28	47,28	47,27	47,32	0,04	0,08	veľmi presná prognóza	47,28	47,27		$y = -0,154ln(x) + 47,892$	0,1347	
400 m [s]	43,74	43,73	43,72	44,13	0,39	0,89	veľmi presná prognóza	43,76	43,75		$y = -0,254ln(x) + 44,649$	0,2598	
800 m [s]	01:42,51	01:42,53	01:42,54	01:41,01	00:01,50	1,46	presná prognóza	01:42,32	01:42,32		$y = -3E-07x + 0,0012$	0,1323	
1500 m [min]	03:28,22	03:28,17	03:28,12	03:29,27	00:01,05	0,50	veľmi presná prognóza	03:28,24	03:28,20		$y = -2E-05ln(x) + 0,0025$	0,4207	
5000 m [min]	12:45,57	12:45,29	12:45,01	12:51,21	00:05,64	0,74	veľmi presná prognóza	12:45,82	12:45,56		$y = -8E-05ln(x) + 0,0091$	0,429	
10000 m [min]	26:34,76	26:34,03	26:33,32	26:56,74	00:21,98	1,38	presná prognóza	26:35,86	26:35,19		$y = -2E-04ln(x) + 0,0192$	0,5506	
Maratón [hod]	2:04:59	2:04:36	2:04:12	2:04:48	0:00:11	0,15	veľmi presná prognóza	2:04:44	2:04:22		$y = -2E-06x + 0,0015$	0,7427	
Hod diskom [m]	71,34	71,43	71,33	71,45	0,11	0,15	veľmi presná prognóza	71,39	71,48		$y = 0,0495x + 69,644$	0,0906	
Hod oštepom [m]	92,07	93,10	94,44	90,37	1,7	1,85	presná prognóza	91,94	92,94		$y = 0,0045x^2 - 0,2081x^2 + 2,7474x + 82,353$	0,5929	
Hod kladivom [m]	83,24	83,23	83,24	80,99	2,25	2,70	blízke očakávanie	83,23	83,24		$y = 0,0001x^3 - 0,0087x^2 + 0,1165x + 83,623$	0,0429	
Skok do výšky [m]	2,39	2,40	2,41	2,36	0,03	1,26	presná prognóza	2,37	2,38		$y = 2E-05x^2 - 0,0014x^2 + 0,0218x + 2,3115$	0,5985	
Skok do diaľky [m]	8,59	8,59	8,59	8,47	0,12	1,40	presná prognóza	8,59	8,59		$y = 2E-05x^2 - 0,0021x^2 + 0,0523x + 8,2853$	0,3244	
Skok o žrdi [m]	6,03	6,05	6,09	5,95	0,08	1,33	presná prognóza	6,01	6,03		$y = 7E-05x^3 - 0,0037x^2 + 0,055x + 5,8095$	0,3413	
Trojskok [m]	17,89	17,90	17,90	17,98	0,09	0,50	veľmi presná prognóza	17,90	17,90		$y = 0,1634ln(x) + 17,306$	0,3651	
Vrh guľou [m]	22,42	22,43	22,44	22,41	0,01	0,04	veľmi presná prognóza	22,43	22,44		$y = 21,934e^{0,008x}$	0,0906	
3 000 m prek. [min]	07:58,58	07:58,42	07:58,28	08:00,90	00:02,32	0,48	veľmi presná prognóza	07:58,60	07:58,45		$y = -5E-05ln(x) + 0,0057$	0,4948	
20 km chôdza [hod]	1:17:44	1:17:41	1:17:39	1:18:24	0:00:40	0,86	veľmi presná prognóza	1:17:48	1:17:46		$y = -8E-06ln(x) + 0,0029$	0,4908	
50 km chôdza [hod]	3:38:26	3:38:21	3:38:17	3:40:37	0:02:11	1,00	veľmi presná prognóza	3:38:01	3:38:37		$y = -1E-05ln(x) + 0,0026$	0,1474	
Štáleta 4x100 m [s]	37,51	37,53	37,52	37,45	0,06	0,16	veľmi presná prognóza	37,49	37,51		$y = -0,0186x + 37,936$	0,2134	
Štáleta 4x400 m [min]	02:56,73	02:56,36	02:56,34	02:58,83	0:02,13	1,21	presná prognóza	02:58,20	02:58,44		$y = 3E-10x^3 + 9E-08x^2 - 3E-06x + 0,0021$	0,1437	
Desaťboj [body]	8617	8560	8496	8483	133,79	1,55	presná prognóza	8560	8496		$y = -0,053x^2 + 1,8122x^2 + 0,4309x + 8333,8$	0,399	

Hodnotenie asociácie DNA markerov k úžitkovým vlastnostiam zvierat pomocou neurónových sietí

Evaluation of DNA markers association in quality traits using neural networks

Ján Buleca, Michal Šoltés, Ladislav Mura, Dagmar Kozelová

Abstract:

Neural network and fuzzy c-cluster analysis have been used for evaluation of relationship between Microsatellite loci and economically relevant traits. We used White Improved swine population as a model. Firstly, we evaluated 10 systems, which defined the major neurons; animals expressed by genotypes have been used as sentences in neural network. Genotype was the input layer for neural network. Neural network was used as a neuron expression tool, according to neurons associated with improvement or degradation of economically relevant traits expressed by computed breeding value. According to high level of variability and potential error, output range of neural network has been restricted to tree answers: improve, degrade, indifferent. Optimality of model has been tested by many neural network structures philosophies as perception, multilevel network, recurrent network, Hopfield's network etc. Assessment to testing ratio of neural network was constructed for the each network architecture separately.

Key words: neural network, fuzzy c-cluster, microsatellite

Kľúčové slová: neurónová sieť, fuzzy c-zhluk, mikrosatelit

1. Úvod

Vzťah genotypu vyjadreného pomocou výsledkov analýzy DNA k úžitkovým vlastnostiam jednotlivých druhov zvierat je základná otázka, ktorou sa zaoberajú ako genetici, tak aj zoolozi. Zistenie priameho vzťahu určitého genotypu na vlastnosť má značné uplatnenie v šľachtení, a ekonomike riadenia. Doteraz sa na vyjadrenie vzťahu medzi genotypom a fenotypom používala a stále používa základná štatistika. Ďalší zo spôsobov, ktorý by sa dal použiť v hodnotení vzťahu medzi genotypom a úžitkovými vlastnosťami je použitie neurónových sietí a využitie fuzzy logiky.

Zadefinovanie neurónovej siete je pomerne zložitá. Neurónová sieť si treba predstaviť ako masívny paralelný výpočtový proces, ktorý má schopnosť informácie uchovávať, spracovávať a upravovať, pričom celý tento systém má snahu napodobniť ľudský mozog a to hlavne v snahe učenia a uchovania informácií, ktoré si vzájomne vymieňajú medzi neurónovými spojeniami – takzvaným synaptickým spojením. Základom neurónovej siete je neurón. Neurón má v sebe niekoľko vstupov a jeden prípadne viac výstupov. Vstupmi dostáva neurón informáciu, ktorá sa spracuje – spracovanie je zväčša jednoduchou lineárnou funkciou. Toto spracovanie je jednoduché, avšak zložitost' tohto systému spočíva v spojení týchto jednoduchých elementov do celku. Pri neurónovej sieti je možné jednotlivé neuróny usporiadať podľa určitých pravidiel - vznikajú takzvané vrstvy. Neurónové siete podľa šírenia informácií môžeme rozdeliť do dvoch základných skupín [8]. Najčastejšie používanými sieťami sú dopredné siete. Informácia sa cez neurónovú sieť šíri len jedným smerom. Druhou skupinou sú neurónové siete rekurentné.

Základnou vlastnosťou neurónovej siete je proces učenia [6]. Proces učenia možno charakterizovať ako obdobie spracovávanie informácie s tým, že sa nastavujú jednotlivé

synaptické váhy. Podľa učenia môžeme neurónové siete rozdeliť tiež do dvoch veľkých skupín. Prvá je neurónová sieť s kontrolovaným učením, druhou skupinou sú neurónové siete s nekontrolovaným učením. Podľa tohto možno rozdeliť neurónové siete do štyroch hlavných skupín. Hlavným predstaviteľom doprednej neurónovej siete s kontrolovaným učením je viacvrstvový perceptrón-MLP, ide o najviac používaný typ neurónovej siete ktorý je použiteľný hlavne pri klasifikácii. Hlavným predstaviteľom doprednej neurónovej siete s nekontrolovaným učením patrí Kohonenove samo organizujúce sa mapy, tieto neurónové siete sa veľmi dobre uplatňujú v procese zhľukovania. Medzi typických predstaviteľov rekurentnej neurónovej siete s kontrolovaným učením patrí Hopfieldová neurónová sieť charakteristická duálnymi neurónmi. Tieto sa používajú hlavne na realizáciu pamäte. Predstaviteľom poslednej skupiny je rekurentná neurónová sieť s nekontrolovaným učením a to ART. Táto sieť predstavuje veľmi veľký potenciál, hlavne z dôvodu reagovania na nové vzory bez toho aby musel nanovo prebehnúť proces učenia. Neurónové siete možno použiť pri aproximácii funkcií, klasifikátoroch a zhľukovaní, predikcii, transformácii, simulácii pamäti a riadení procesov.

V oblasti použitia neurónových sietí v kombinácii s mikrosatelitnými markermi bolo publikovaných pomerne málo prác. Väčšina prác je zameraná na hodnotenie diverzity pomocou zhľukových analýz, alebo využívania Bayesiansej logiky pri hodnotení premiešania populácií [2, 4, 10, 11].

Autori [1] zamerali svoj výskum na získanie nástroja pre odlišenie prirodzeného a umelo vyľahnutého pstruha (*Salmo trutta*, L.) v Atlantickej populácii. Podobná práca zameraná všeobecne na identifikáciu rýb použitím kombinácie mikrosatelitov a neurónových sietí bola publikovaná kolektívom odborníkov [3].

2. Materiál a metódy

V našom modelovom prípade sme sa snažili pomocou neurónovej siete nájsť mikrosatelitný systém, prípadne systémy, ktoré majú vplyv na zlepšenie, prípadne zhoršenie hrúbky slaniny u ošípanej. U plemena Slovenská biela ušľachtilá sme na vzorke 246 jedincov publikovanej v práci [11] sledovali vzťah medzi desiatimi mikrosatelitnými systémami (PSW24, PSO107, PSO068, PSW936, PSW353, PSO386, PSO355, PSW72, PTNFB, PSO070) a úžitkovou vlastnosťou vyjadrenou pomocou vypočítanej plemennej hodnoty hrúbky slaniny [5]. Pre zjednodušenie celého modelu sme reálne hodnoty hrúbky slaniny previedli do formátu indierentný, zlepšovateľ, zhoršovateľ. Transformáciu sme vykonali pomocou fuzzy klasifikátorov. Každému zvieraťu sme priradili číslo 1 ak sa jeho plemenná hodnota hrúbka slaniny nachádzala v intervale zlepšovateľov, číslo 0 ak sa jeho plemenná hodnota hrúbka slaniny nachádzala v intervale indierentný a číslo 2 ak sa jeho plemenná hodnota hrúbka slaniny nachádzala v intervale zhoršovateľ.

Do neurónovej siete sme zvolili len také jedince, kde mikrosatelitná alela bola detegovaná. Jednotlivé mikrosatelitné lokusy sme zadefinovali vo vstupnej vrstve neurónovej siete ako jednotlivé neuróny. Neurón pritom dostával informáciu v podobe dvoch číselných informácií (dve alely). Zdefinovali sme jeden neurón na výstupe, ktorý bol definovaný prevedenou hodnotou hrúbky slaniny. Výstup z tohto neurónu bol v podobe jedného čísla. Počet viet použitých v neurónových sieťach bol 2368. Spracovanie prebiehalo pomocou softwaru MATLAB [9] a STATISTICA 6.1 [7]. V procese analyzovania vhodnej neurónovej siete sme použili 426 alternatív neurónových sietí.

3. Výsledky a diskusia

Za účelom zistenia vhodného analytického postupu sme použili na modelovej populácii ošpaných všetky štyri základné skupiny neurónových sietí. V rámci dopredných neurónových sietí s kontrolovaným učením - MLP sme použili siete s 10 neurónmi na vstupe, ktoré predstavovali 10 mikrosatelitných lokusov. Postupne bolo preskúšaných viacero alternatív neurónových sietí s rôznym počtom skrytých vrstiev (1-10), pričom pre každú skrytú vrstvu bol rôzne menený počet neurónov. Výstupný neurón bol vždy len jeden. Klasifikačný pomer (Tabuľka 1) poukazuje na nejednoznačnosť použitia neurónovej pri súčasnom nadefinovaní neurónov pre zvolený biologický model.

Tabuľka 1 Prehľad použitých neurónových sietí na modelovej skupine zvierat

Typ neurónovej siete	Typ neurónovej siete	Počet neurónov na vstupe	Počet skrytých vrstiev	Počet výstupných neurónov	Klasifikačný pomer
Dopredná NS s kontrolovaným učením MLP	10-X-1	10	1	1	0.23-0.36
	10-X-X-1	10	2	1	0.11-0.16
	10-X-X-X-1	10	3	1	0.26-0.363
	10-X-X-X-X-1	10	4	1	0.24-0.362
	10-X...X-1	10	5- 10	1	0.26-0.31
Dopredná NS s nekontrolovaným učením	Kohonenova	10	0	1	0.245-0.246
Rekurentná NS s kontrolovaným učením	Hopfieldová		Počet duálnych neurónov 10		0.11-0.115
Rekurentná NS s nekontrolovaným učením	ART	10	1- 36	1	0.22-0.314

Zdroj: vlastné spracovanie

Pri použití doprednej neurónovej siete s nekontrolovaným učením (Kohonenova sieť) sme pri použití 10 neurónov na vstupe a 1 na výstupe pri použití rôznych algoritmov dosiahli nižšie hodnoty klasifikačného pomeru (0.245-0.246). Ešte nižšie hodnoty klasifikačného pomeru (0.11 – 0.115) boli vypočítané použitím rekurentnej neurónovej siete s kontrolovaným učením (Hopfieldova sieť). Rekurentná neurónová sieť s nekontrolovaným učením (ART) bola použitá hlavne pre jej vhodnosť pri analyzovaní biologického materiálu. Klasifikačný pomer pre tento typ siete bol rovnako nízky (0.22-0.314), ako u ostatných použitých sietí.

Výsledky prezentované v tabuľke 1 naznačujú, že definovanie základných neurónov v podobe jednotlivých mikrosatelitných lokusov nie je dostatočné pre zabezpečenie vysokého klasifikačného pomeru. Príčinou je pravdepodobne značná variabilita v jednotlivých lokusov a aj nejasný vzťah medzi úžitkovosťou a genetickým markerom.

Ako najvhodnejší variant neurónovej siete s kontrolovaným učením – MLT sa javil typ siete 10-12-14-11-1 pri klasifikačnom pomere – 0.363. Základné charakteristiky tejto neurónovej siete sú popísané v tabuľke 2.

Tabuľka 2 Základné charakteristiky najoptimálnejšej neurónovej siete

charakteristika	Popis	
Typ siete	Dopredná viacvrstvová sieť s kontrolovaným učením – perceptrón – MLP typ 10-12-14-11-1	
Rozloženie a normalita váh okolo 0	Test preukázal symetriu váh okolo 0, normalita synaptických váh – bola preukazná na $\alpha=0,05$	
Testovanie neurónovej siete Na testovanie neurónovej siete bolo použitých náhodných 828 viet čo predstavuje približne 35 % z celkového počtu viet – vyššie percento by mohlo priniesť pretrénovanie neurónovej siete	Trénovacia množina	262 viet , klasifikačný pomer 0,344
	Trénovacia testovacia množina	252 viet, klasifikačný pomer 0,362
	Validation test (overovací test)	314 viet, klasifikačný pomer 0,363
Fáza žitia	Počet viet 1540	Percento správne klasifikovaných jedincov 36 %

Zdroj: vlastné spracovanie

Analýzu získaných výsledkov sme kontrolovali iným metodickým prístupom, ktorý predpokladal v prípade výrazného vzťahu mikrosatelitného lokusu k úžitkovej vlastnosti zvýšenie stupňa príslušnosti k jednému zo zhlukov. Predpokladáme, že pri definovaní jednotlivých alel pozorovaných lokusov ako vstupných neurónov by bolo možné získať vyššie hodnoty klasifikačného pomeru.

4. Záver

Napriek tomu, že nám sa nepodarilo jednoznačne definovať vhodnú neurónovú sieť, myslíme si, že použitie neurónových sietí v genetike je perspektívne. Za najlepšiu sieť sa nám zdala sieť MLP 10-12-14-11-1 s klasifikačným pomerom 0.363. Tento pomer je však značne nízky a preto nemožno jednoznačne definovať jednotlivé mikrosatelitné lokusy, ktoré majú priamy vplyv na úžitkovosť. Záverom môžeme konštatovať že analyzovanie vzťahu medzi úžitkovými vlastnosťami a genotypom nie je vhodné analyzovať postupom, v ktorom sa ako vstupné neuróny zvolia celé mikrosatelitné lokusy. Riešenie vidíme v zameraní sa na jednotlivé formy génov (alely) a ich použitie ako vstupných neurónov.

5. Literatúra

- [1] AURELLE, D., LEK, S., GIRAUDEL, J-L. and BERREBI, P. 1999. Microsatellites and artificial neural networks: tools for the discrimination between natural and hatchery brown trout (*Salmo trutta*, L.) in Atlantic populations. In *Ecological Modelling*. 120(2-3):313-324.

- [2] BURÓCZIOVÁ, M., ŘÍHA, J., ŽIDEK, R., TRANDŽÍK, J. and JAKABOVÁ, D. 2008. Genetic structure of nine horse populations. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 56(2), 57-60.
- [3] HANSEN, M. M., KENCHINGTON, E., NIELSEN, E. E. 2001. Assigning individual fish to populations using microsatellite DNA markers. In *Fish and Fisheries*. 2(2):93-112
- [4] MARŠÁLKOVÁ, L., ŽIDEK, R. and POKORÁDI, J. 2010. Genetic and statistical analysis of founder lines in red deer population. *Acta fytotechnica et zootechnica*, 13(1), 24-28. Slovenská poľnohospodárska univerzita v Nitre. Retrieved January 21, 2011, Dostupné na: [http://www.fem.uniag.sk/acta/sk/1/acta_fytpotechnica_et_zootechnica_\(online\)/obsah/2010/1/839/](http://www.fem.uniag.sk/acta/sk/1/acta_fytpotechnica_et_zootechnica_(online)/obsah/2010/1/839/).
- [5] PEŠKOVIČOVÁ, D., WOLF, J., GROENEVELD, E., WOLFOVÁ, M. 2002. Simultaneous estimation of the covariance structure of traits from field test, station test and litter recording in pigs. In: *Livestock production science*. vol. 77, 2002, iss. 2-3, p. 155-165.
- [6] SIMA, J. 1998. Introductions to Neural Networks. Technical report No V-755, ICS CAS, Prague, 1998.
- [7] STATISTICA 6.1, 2007. Neural Networks. Electronic Statistics Textbook. StatSoft, Inc., Tulsa, OK, USA. Dostupné na: <http://www.statsoft.com/textbook/stathome.html>
- [8] STEHLÍKOVÁ, B. 2004. Nový metodologický prístup k prognózovaniu demografických časových radov. In: *Acta oeconomica et informatica*, 2/2004, pp. 49-52. Dostupné na: www.fem.uniag.sk/acta/download.php?id=89
- [9] TOH, K., TODD, M. AND TUTUNCU, R. 1999. Sdpt3 a matlab software package for semidefinite programming. *Optimization Methods and Software*. 1999. 11:545-581
- [10] ZIDEK, R. - KASARDA, R. 2010. Distribution of genetic distance within groups with different relationship coefficient. In: *Acta fytotechnica et zootechnica*, 13(specia issue), 73-76. Slovenská poľnohospodárska univerzita v Nitre. Retrieved April 19, 2011, Dostupné na: <http://www.fem.uniag.sk/acta/download.php?id=821>
- [11] ZIDEK, R., POKORADI, J. and BANDRY, L. 2008. Biodiversity in der microsatellite markers observed by. In: *Journal of Agrobiologia*, 25(1), pp. 113-115.

Pod'akovanie

Predkladaný článok vznikol za podpory Agentúry pre podporu výskumu a vývoja pod číslom APVV 20-063205 a prostredníctvom grantu VEGA č. 1/0046/10.

Adresa autorov:

MVDr. Ing. Ján Buleca, PhD.
Katedra financií
Ekonomická fakulta
Technickej univerzity v Košiciach,
Němcovej 32, 040 01 Košice
E-mail: buleca@gmail.com

Ing. et Bc. Ladislav Mura, Ph.D.
Ústav ekonomie
Moravská vysoká škola Olomouc
Jeremenkova 42
772 00 Olomouc, Česká republika
E-mail: ladislav.mura@gmail.com

Ing. Michal Šoltés, PhD.
Katedra bankovníctva a investovania
Ekonomická fakulta
Technickej univerzity v Košiciach,
Němcovej 32, 040 01 Košice

Ing. Dagmar Kozelová, PhD.
Katedra hygieny a bezpečnosti potravín
Fakulta biotechnológií a potravinárstva
Slovenská poľnohospodárska univerzita v Nitre
Trieda A. Hlinku 2, 949 76 Nitra
E-mail: dkozelova@gmail.com

Názory študentov na vyučovanie matematiky a prírodovedných predmetov Academic opinion of mathematical and science education

Soňa Fándlyová, Lubomír Rybanský

Abstract: The thesis deals with the mathematic knowledge into science by the help of questionnaire. We use program STATISTICA by the evaluation of questionnaire. The aim of research is interest of mathematics and science and its seriousness in school education.

Key words: statistics, mathematics, sciences

Kľúčové slová: štatistika, matematika, prírodné vedy

JEL classification: ZOO.

1. Metodika a ciele dotazníka

Jedným z cieľom našej práce bolo zistiť názory študentov na vyučovanie prírodovedných predmetov formou dotazníka, ktorý obsahoval osem otázok s možnosťou výberu odpovede. Zaujímal nás názor študentov na zaujímavosť a náročnosť vybraných prírodovedných predmetov a matematiky. Špeciálne sme sa zamerali na matematiku a chémiu, či učiteľ dokáže študentov zaujať a aký vplyv má záujem o predmet s náročnosťou predmetu a naopak.

Dotazník sme realizovali na Piaristickom gymnáziu sv. J. Kalazanského v Nitre. Zúčastnilo sa ho 148 respondentov vyššieho gymnázia zo siedmich tried.

Cieľom dotazníka bolo zistiť u študentov:

- mieru záujmu o predmet chémiu
- mieru záujmu o predmet matematika
- mieru náročnosti predmetu chémiu
- mieru náročnosti predmetu matematika
- porovnanie záujmu o predmety medzi matematikou a prírodnými vedami
- porovnanie náročnosti matematiky a prírodovedných predmetov z hľadiska prípravy na vyučovanie

2. Obsah a vyhodnotenie dotazníka

V otázke číslo 3 sme dali študentom možnosť vybrať si z dvojice uvedených prírodovedných predmetov ten, ktorý je pre nich zaujímavejší. V otázke číslo 6 sme im z tých istých dvojíc požiadali o výber toho predmetu, ktorý je náročnejší z hľadiska prípravy na vyučovaciu hodinu. Na základe odpovedí na uvedené otázky sme chceli zistiť, či a akým spôsobom závisí zaujímavosť predmetu od jeho náročnosti.

Testujeme výskumnú hypotézu:

H_0 : *Zaujímavosť predmetu nezávisí od náročnosti prípravy na tento predmet.*

(Hypotézu H_0 chceme zamietnuť a tým nadobúda platnosť alternatívna hypotéza H_A)

H_A : *Zaujímavosť predmetu závisí od náročnosti prípravy na tento predmet.*

Pri každej dvojici predmetov sledujeme hodnoty premennej A – *náročnosť prípravy* a premennej B – *zaujímavosť predmetu*. Obe premenné sú kategoriálne.

Nech p_i , resp. p_j je pravdepodobnosť, že premenná A (náročnosť) nadobudne hodnotu i resp. premenná B (zaujímavosť) nadobudne hodnotu j . V prípade, že premenné A a B sú nezávislé, tak pravdepodobnosť p_{ij} , že u daného objektu zistíme hodnotu premennej A rovnú i a hodnotu premennej B rovnú j je súčinom pravdepodobností p_i a p_j .

Testujeme štatistickú hypotézu nezávislosti

$$H_0: p_{ij} = p_i \cdot p_j, \quad i = 1, 2, j = 1, 2.$$

Testovacou štatistikou pri posudzovaní nezávislosti v prípade štvorpoľnej tabuľky je

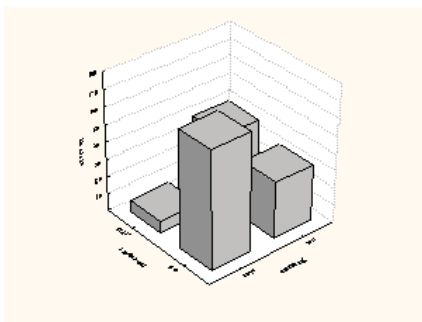
$$\chi^2 = n \sum_{i=1}^2 \sum_{j=1}^2 \frac{n_{ij}^2}{n_i \cdot n_j} - n = n \frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_1 \cdot n_2 \cdot n_1 \cdot n_2},$$

ktorej hodnotu porovnáваме s kritickou hodnotou $\chi_1^2(0,05) = 3,84$.

Po vykonaní náhodného výberu s rozsahom 148 a zistení príslušných početností sme vytvorili kontingenčné tabuľky pre testovanie závislosti dvoch premenných u všetkých dvojíc skúmaných predmetov. Výsledky sú v tabuľkách 1 až 7.

Tabuľka 1: MAT – BIO

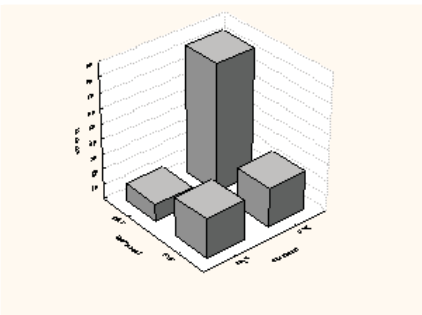
A náročnosť	B - záujem		riadkové súčty
	MAT	BIO	
MAT	7	68	75
BIO	40	33	73
stĺpcové súčty	47	101	148



Obrázok 1: Mat- Bio

Tabuľka 2: MAT – FYZ

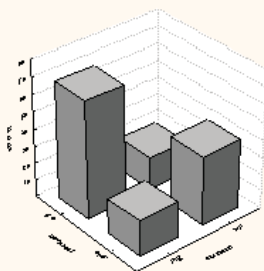
A náročnosť	B - záujem		riadkové súčty
	MAT	FYZ	
MAT	11	26	37
FYZ	85	26	111
stĺpcové súčty	96	52	148



Obrázok 2: Mat-Fyz

Tabuľka 3: BIO – CHE

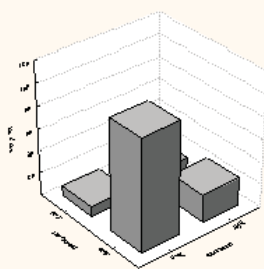
A náročnosť	B – záujem		riadkové súčty
	BIO	CHE	
BIO	19	40	59
CHE	68	21	89
stĺpcové súčty	87	61	148



Obrázok 3: Bio-Che

Tabuľka 4: CHE – FYZ

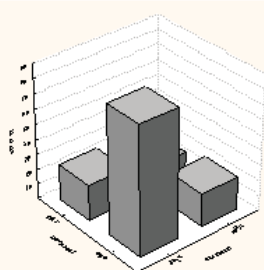
A náročnosť	B – záujem		riadkové súčty
	CHE	FYZ	
CHE	25	13	38
FYZ	101	9	110
stĺpcové súčty	126	22	148



Obrázok 4: Che-Fyz

Tabuľka 5: MAT – GEO

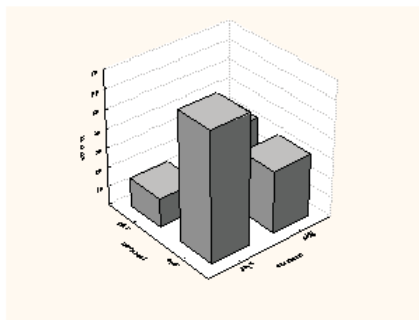
A náročnosť	B – záujem		riadkové súčty
	MAT	GEO	
MAT	24	86	110
GEO	16	27	43
stĺpcové súčty	40	113	148



Obrázok 5: Mat-Geo

Tabuľka 6: MAT – CHE

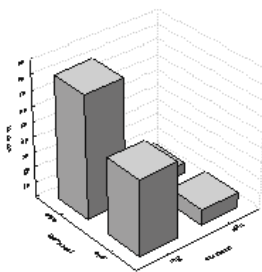
A náročnosť	B - záujem		riadkové súčty
	MAT	CHE	
MAT	15	67	82
CHE	34	32	66
stĺpcové súčty	49	99	148



Obrázok 6: Mat-Che

Tabuľka 7: CHE – GEO

A náročnosť	B - záujem		riadkové súčty
	CHE	GEO	
CHE	50	81	131
GEO	10	7	17
stĺpcové súčty	60	88	148



Obrázok 7: Che-Geo

Vypočítali sme hodnoty testovej štatistiky χ^2 , porovnali ju s kritickou hodnotou a taktiež uvádzame príslušnú p – hodnotu. Výsledky sú v tabuľke 8.

Tabuľka 8: χ^2 test nezávislosti v kontingenčných tabuľkách

dvojica	χ^2	kritická. hodnota	p - hodnota
MAT-BIO	35,28	3,84	0,000
BIO-CHE	28,61	3,84	0,000
MAT-FYZ	26,72	3,84	0,000
CHE-FYZ	15,12	3,84	0,000
MAT-GEO	5,89	3,84	0,015
CHE-GEO	2,66	3,84	0,102
MAT-CHE	18,22	3,84	0,000

S výnimkou dvojice predmetov CHE – GEO u všetkých ostatných dvojíc predmetov hypotézu o nezávislosti premenných *zaujímavosť predmetu* a *náročnosť prípravy* na hladine významnosti 0,05 zamietame, nakoľko dosiahnuté p – hodnoty sú menšie ako 0,05. Neplatnosť štatistickej hypotézy znamená aj neplatnosť výskumnej hypotézy.

Navyše z hodnôt v kontingenčných tabuľkách vyplýva, že ten z dvojice predmetov, ktorý si vyžaduje menej náročnú prípravu je vnímaný ako zaujímavejší. Samozrejme toto tvrdenie platí aj obrátene, teda ak je predmet zaujímavejší, tak je menej náročný na prípravu

Z kontingenčných tabuliek 1–7 sme vytvorili tabuľku 9, v ktorej sú početnosti obľúbenejšieho resp. náročnejšieho predmetu z uvedenej dvojice spolu s percentuálnym podielom obľúbenejšieho resp. náročnejšieho predmetu.

Tabuľka 9: Početnosti obľúbenejšieho resp. náročnejšieho predmetu pre každú z dvojíc predmetov

		záujem			náročnosť		
predmet 1	predmet 2	predmet 1	predmet 2	zaujímavejší v %	predmet 1	predmet 2	náročnejší v %
MAT	BIO	47	101	68,24	75	73	50,68
BIO	CHE	87	61	58,78	59	88	59,86
MAT	FYZ	96	52	64,86	37	111	75,00
CHE	FYZ	126	22	85,14	38	110	74,32
MAT	GEO	40	108	72,97	110	38	74,32
CHE	GEO	60	88	59,46	131	17	88,51
MAT	CHE	49	99	66,89	82	66	55,41

Predmet *matematika* je vnímaný ako zaujímavejší predmet iba v porovnaní s *fyzikou* (64,86%) v porovnaní s *chémiou* má podiel (33,11%), *biológiou* (31,76%), *geografiou* (27,03%)

Z hľadiska náročnosti prípravy na vyučovanie je vnímaný ako náročnejší v porovnaní s *geografiou* (74,32%), *chémiou* (55,41%) s *biológiou* je približne rovnako náročný a menej náročný je v porovnaní s *fyzikou* (25%).

Predmet *chémia* je pre študentov zaujímavejší v porovnaní s *fyzikou* (85,14%) a *matematikou* (66,89%). Menej zaujímavým je v porovnaní s *biológiou* (41,22%) a *geografiou* (40,54%).

Chémia je považovaná za náročnejší predmet v porovnaní s *geografiou* (88,51%) a *biológiou* (59,86). Menej náročná je v porovnaní s *matematikou* (44,49%) a *fyzikou* (25,68%).

Na pomyselnom rebríčku obľúbenosti piatich prírodovedných predmetov by sa *matematika* umiestnila na štvrtom mieste a *chémia* na treťom mieste. V rebríčku náročnosti predmetov by *matematike* patrila druhá priečka a *chémii* tretia priečka.

V tabuľke 10 uvádzame počty študentov u ktorých je predmet *matematika* (*chémia*) zaujímavejší resp. náročnejší v porovnaní s ostatnými prírodovednými predmetmi.

Tabuľka 10: Náročnosť resp. väčšia zaujímavosť predmetov *matematika* a *chémia* voči ostatným prírodovedným predmetom

		záujem		náročnosť			
predmet		predmet	ostatné	predmet %	predmet	ostatné	predmet %
MAT	ostatné	18	130	12,16	24	124	16,22
CHE	ostatné	28	120	18,92	16	132	10,81

Vzťah medzi mierou záujmu

Zaujímalo nás, či existuje vzťah medzi záujmom o daný predmet, náročnosťou predmetu vzhľadom na prípravu a tým ako dokáže učiteľ žiaka zaujať na vyučovacej hodine. Uvedené závislosti sme sa rozhodli skúmať u predmetov matematika a chémia.

Matematika

V rámci dotazníka mali študenti odpovedať na otázky

o2 - Vyznačte mieru vášho záujmu o predmet matematika.

o5 - Vyznačte mieru náročnosti predmetu matematika

o8 - Vyznačte mieru akou Vás učiteľ dokáže zaujať na hodine matematiky.

Škála odpovedí bola od 1 do 5, pričom 1 vyjadruje najnižšiu mieru a 5 najvyššiu mieru.

Po spracovaní dotazníka sme získali výsledky, ktoré uvádzame v tabuľke 7.11.

Tabuľka 11: Výsledky odpovedí na otázky o2, o5, o8

otázka	kódy - početnosti					priemer	kódy - percentuálny podiel				
	1	2	3	4	5		1	2	3	4	5
o2	31	41	40	23	13	2,64	20,95	27,70	27,03	15,54	8,78
o5	14	27	37	41	29	3,30	9,46	18,24	25,00	27,70	19,59
o8	63	29	24	20	12	2,25	42,57	19,59	16,22	13,51	8,11

Nech premenná X_2 vyjadruje odpoveď na otázku o2 a môže nadobúdať hodnoty 1, 2, 3, 4, 5. Podobne definujeme premennú X_5 – odpoveď na otázku o5 a premennú X_8 odpoveď na otázku o8 u každého z opýtaných študentov.

Všetky uvedené premenné sú kategoriálne ordinálne premenné, pretože rozdiel medzi jednotlivými úrovňami chápeme kvantitatívne. Na posúdenie závislosti medzi premennými použijeme Goodman – Kruskalov koeficient korelácie γ a Kendallov koeficient korelácie τ .

Hodnoty koeficientov korelácie medzi dvojicami premenných spolu s ich významnosťou sú v tabuľke 12. Nad diagonálou sú hodnoty koeficientu korelácie γ a pod diagonálou hodnoty koeficientu korelácie τ .

Tabuľka 12: Koeficienty korelácií medzi odpoveďami na otázky o2, o5 a o8.

	o2	o5	o8
o2		-0,304	0,510
o5	-0,242		-0,283
o8	0,395	-0,217	

Červenou farbou sú vyznačené štatisticky významné závislosti na hladine významnosti 0,05.

Z výsledkov vyplýva, že:

- čím je u študenta menšia miera záujmu o predmet matematiky, tým je tento predmet náročnejší ($\gamma = -0,304$, $\tau = -0,242$)

- čím viac dokáže vyučujúci na hodine matematiky žiaka zaujať, tým je o predmet väčší záujem ($\gamma = 0,510$, $\tau = 0,395$)

- čím viac dokáže vyučujúci na hodine matematiky žiakov zaujať, tým je príprava na hodiny matematika menej náročná ($\gamma = -0,283$, $\tau = -0,217$).

Chémia

Analogické otázky ako v prípade predmetu matematika, dostali študenti aj ohľadom predmetu chémia.

o1 - Vyznačte mieru vášho záujmu o predmet chémia.

o4 - Vyznačte mieru náročnosti predmetu chémia.

o7 - Vyznačte mieru akou Vás učiteľ dokáže zaujať na hodine chémia.

Pri vyhodnocovaní sme postupovali rovnako ako v prípade predmetu matematika. Výsledky sú uvedené v tabuľke 13.

Tabuľka 13: Koeficienty korelácií medzi odpoveďami na otázky o1, o4 a o7.

	o1	o4	o7
o1		-0,077	0,462
o4	-0,051		-0,037
o7	0,335	-0,031	

Červenou farbou sú vyznačené štatisticky významné závislosti. Z odpovedí na uvedené tri otázky vyplýva, že

- miera záujmu o predmet chémia nezávisí od náročnosti tohto predmetu ($\gamma = -0,077$, $\tau = -0,051$)
- miera záujmu o predmet závisí od toho ako dokáže vyučujúci zaujať na hodine a to tak, že čím viac dokáže zaujať, tým je o predmet väčší záujem ($\gamma = -0,462$, $\tau = -0,335$)
- miera náročnosti predmetu nezávisí od toho ako dokáže vyučujúci zaujať na hodine chémie ($\gamma = -0,037$, $\tau = -0,031$).

3. Závěry výskumu

V prvom rade by sme mali podporiť význam matematiky v chémii. Znovu objaviť prelínanie matematiky s prírodnými vedami, a zasvätiť žiakov do „skrytých“ tajomstiev prírodných vied. Výchovno-vzdelávací proces bude ucelený, ak žiacividia súvislosti medzi predpísanými pravidlami v matematike a následnej aplikácii v chémii. Je dôležité naučiť žiakov logicky myslieť nielen na hodinách matematiky. Ťažká, nie však nezládnuteľná úloha. Tým, že sa žiaci naučia aplikovať matematické vedomosti aj na iných hodinách, nebude sa im matematika javiť ako zbytočná. Vytvárať u žiakov tvorivý prístup k práci a vniesť kreatívny prístup k získavaniu nových poznatkov a zákonitostí. Práve riešenie problémov, ktoré nemajú na prvý pohľad priamy súvis s matematikou, je tá najťažšia časť matematického vzdelávania. Žiaci nemajú prax v takomto procese a potencionálna schopnosť matematiky pomôcť im pri riešení problémov, s ktorými sa stretnú, nemusí byť plne realizovaná. Medzi najdôležitejšie zložky vo vzdelávaní patrí rozvíjanie predstavivosti a tvorivosti žiakov. Zastávame názor, že vyvrcholením tvorivej činnosti je zážitok z poznania niečoho nového.

4. Literatúra

- [1]FÁNDLYOVÁ, S. 2010. Matematika v prírodovedných predmetoch. Dizertačná práca.
 [2]HENDL, J. 2009. Přehled statistických metod. Portál 2009. Praha. ISBN 978-80-7367-482-3

Adresa autorov:

Soňa Fándlyová, PaedDr., PhD.,
 Tr. A. Hlinku 2
 949 76 Nitra
 sona.fandylova@fem.uniag.sk

Lubomír Rybanský, RNDr.
 Tr. A. Hlinku 1
 949 74 Nitra
 lubomir.rybansky@ukf.sk

**Souvislost instrumentálních charakteristik sýrů zrajících pod mazem
s délkou zrání**
**Relation of instrumental characteristics of acid smeared cheeses to ripening
time**

David Hampel, Šárka Hanzelková, Aneta Bubeníčková

Abstract: Acid curd smeared cheese was ripened for 35 days and the ripening time was assessed as a main factor of development of the colour and the texture of cheese. Hardness was measured using shear device Tira test 27025 (ball penetration probe), objective measurement of colour was provided using a spectrophotometer Konica Minolta CM 3500d. Analysis of variance at 5% significance level showed that both colour and texture were influenced by the ripening time. The texture improved at the end of shelf life of the cheese, hardness decreased by 45 %. The most valuable parameter of colour was the lightness, which decreased significantly ($p < 0.001$) during ripening. The hue considerably changed and was unified in matured cheese. There was interesting significant positive correlation between hardness and lightness (cut 0.8447, surface 0.8606, $p < 0.001$).

Key words: Cheese, Correlation, General Linear Model, Hardness, Hue, Lightness, Spectrophotometer

Klíčové slová: Cheese, Correlation, General Linear Model, Hardness, Hue, Lightness, Spectrophotometer

JEL classification: C31

1. Introduction

Acid curd smeared cheese is a typical food that acquire its ideal characteristics only after specific aging time. These characteristics are determinant for consumers, who decide on the success of the product. Before the consummation, when taste and smell of the cheese is assessed, consumers judge the texture (hardness) and colour of packed cheese during shopping. The expectations cannot be fulfilled if the cheese is insufficiently ripened and hard and the coverage by the colourful smear culture is deficient. Measurement of the colour and the texture can be also useful for product quality control.

Primary ripening takes place at the temperature from 25 to 30 °C with the contribution by homofermentative Lactococcus, which is followed by the smear culture - especially Brevibacterium linens, Arthrobacter and other coryneform bacteria and several species of yeasts (Geotrichum candidum, Kluyveromyces marxianus, Debaryomyces hansenii) [7]. After deacidifying of the surface of the cheese the smear culture is applied and at the temperature of 15 °C important changes in colour, taste, aroma and texture take place in several days. Typical dark yellow, orange, pink or red-brick rind's colour originates in the synthesis of carotenoids and other pigments by culture microorganisms [4], essential is Brevibacterium linens [1]. Ripening of the cheeses made from acid curd is typical by the visible change of white acid curd into a yellowish-brown cheese with smooth and elastic texture which proceeds from the surface to the core [2].

Colour and texture can be evaluated by the sensory analysis (by human senses) as well as by the modern instrumental methods. The advantages of these methods are accuracy, repeatability, possibility of more detailed statistical analysis and lower time consumption. Many authors have been concerned with the texture of cheeses [3] as well as with the changes during ripening of smeared cheeses [8]. The evaluation of the colour of cheese by spectrophotometry has not been used very often [4].

The aim of the study was to describe the characteristics of the acid smear cheese by the instrumental measurement of colour and hardness during ripening. The relationship between both instrumental methods was statistically evaluated.

2. Material, methods

The cheese was sampled in regular intervals after the production: 7th day after drying (T4), 10th day after ripening (T5) and 38th day at the expiry date (T6). Always six packages of cheese were investigated. The samples were ripened at temperature of 7 °C. A shear device Tira test 27025 was used for the instrumental evaluation of texture. The hardness was measured by the penetration test (hardness was determined as the maximum of force (N) needed for the penetration of the ball probe into the cheese surface). Objective measurement of colour on the surface and on the cut of the cheese samples was provided using a spectrophotometer Konica Minolta CM 3500d. The optical system is using diffused illumination and reflected light was measured under coal 8 (d/8) with SCE function (specular component excluded) for elimination of mirror radiance. Slot diameter was 8 mm. All spectrum between 380-780 nm was measured and colour was numerically defined in L*a*b* colour system (CIE Lab). In this system value L* is lightness and value a* and b* are axis of colour. +a* is red direction, -a* is green direction, +b* is yellow direction and -b* is blue direction. We obtain totally 90 balanced observations for hardness, 162 for lightness and hue on the surface and 324 for lightness and hue on the cut.

Results were statistically analysed by the means of statistical system Genstat 13.1 (<http://www.vsnl.co.uk/software/genstat/>). General linear model (for example Rao et al. [6]) was used for this purpose. We have two factors, ripening time as the factor we focus on and cheese package as a sampling factor. Analysis of variance followed by pairwise comparison of effects (using predictions and the least significant differences, LSDs) was calculated for all examined characteristics.

Anticipated dependences among measured characteristics were investigated using sample correlation coefficient and appropriate test of independence. Situation is complicated by unequal numbers of observations for particular characteristics. This makes impossible direct calculation of sample correlation coefficients. We solve this problem by using predictions for both factors simultaneously. We get 18 values per characteristic to compare by this manner. It should be noted that presence of factors can lead to very distant groups of data, which can significantly cause sample correlation coefficients in the sense of a false dependence, but this is not our case.

3. Results and discussion

Assumptions for two-way ANOVA analysis (normality of data and equality of variances) were followed in every analysis, which caused only one LSD for comparison of all factor effect predictions. ANOVA-tables (for example see Table 1) were produced for particular characteristics. In all cases, both factors are statistically significant at the significance level 0.05 ($p < 0.001$). It is evident that factor of cheese package lowers model variability and increases distinguishability of the time factor effects. Effects of cheese

package factor are adjusted in predictions of time factor. Now we will focus different characteristics gradually.

Table 1: ANOVA-table for hardness of the cheese with factors Package and Time. Both factors are statistically significant at the significance level 0.05.

Factor	d. f.	s. s.	m. s.	F	F pr.
Package	5	113.296	22.659	10.03	<.001
Time	2	432.629	216.314	95.80	<.001
Residual	82	185.161	2.258		
Total	89	731.086	8.214		

The results of hardness measurement are presented in Table 2. The least significant difference was 0.616 and it is possible to conclude that difference in hardness is statistically significant for all ripening stages. The results imply visible decrease of hardness after 28 days of ripening, respectively at on the expiry date. During the initial ripening phases the cheese mass is harder, the ripening is slow and substantial decrease in hardness appears only after complete development of the smear culture and subsequent deepening of proteolysis.

Table 2: Prediction of hardness of the cheese at the beginning of ripening (T4), during ripening (T5) and at the end of ripening (T6).

	T4	T5	T6
Hardness [N]	11.51	10.32	6.38

Our results are in agreement with [3] who also stated the decrease in hardness of cheeses during ripening and indicated that ripening time from 3 to 4 weeks are optimal for obtaining properties anticipated by consumers. Also [8] evaluated cheese hardness (sensory analysis) and proved best texture properties of cheese after ripening for 4 weeks, whereas usually the cheese is placed on the market after approximately only 3 weeks.

Basic parameter of colour characteristics of cheese was the lightness measured on the surface and on the cut of cheese. Predicted values are shown in Table 3. After classifying the samples according the ripening phase into initial stage with white curd in the core of the cheese and final stage (cheese is ripened and coloured also in the centre of cheese mass) the decrease of lightness values is considerable. Statistically significant ($p < 0.001$) differences were found among the ripening stages (Table 3). The least significant difference for this analysis was 0.781 for the lightness on the cut and 0.393 on the surface.

Table 3: Predicted lightness L^* in three ripening phases (T4, T5, T6) measured on the surface and on the cut of the cheese.

Lightness L^*	T4	T5	T6
surface	76.17	68.96	50.20
cut	91.15	89.98	54.16

Characteristic area of radiance for the smear cheese is area of yellow and red light. Values a^* , b^* which are on the axis of colour in the diagram did not differ considerably among ripening stages despite the fact that colour axis b^* slightly moved from light yellow to

dark yellow colour. It is more appropriate to assess hue h which is an angle on the hue circle ranging from $0-360^\circ$ (hue angle 0° represents red colour). The hue of cheeses during ripening was changing during ripening (Table 4) and the differences among stages were statistically significant ($p < 0.001$) with the least significant difference 0.777 for measurement on the cut and 0.453 for the surface measurement.

The biggest differences between the surface and the core of the cheese were observed in insufficiently ripened cheese. In the final ripening stage the hue of whole cheese mass was unified. Between the lightness and hue on the cheese cut was found statistically significant ($p < 0.001$) correlation -0.7301 and the lightness and hue on the cheese surface also significantly negatively correlate (coefficient -0.8502). These sample correlation coefficients indicate that with decreasing lightness increases hue value, colour of the cheese is entirely changing and is characteristic for a specific ripening phase.

Table 4: Predicted values of hue of the ripening cheese in dependence on the ripening time.

Hue h [°]	T4	T5	T6
surface	89.47	91.82	101.94
cut	95.95	97.37	103.76

Our results are consistent with [4] whose tests resulted in conclusion that in red-smear cheeses dominates orange colour and also yellow colour is rapidly developed. Dufosse et al. [4] also indicated that red colour was increasing gradually during all ripening time. By contraries Galaup et al. [5] claimed that red colour in ripened cheese is rather stable what was approved also in our experiments.

Both parameters – colour and hardness – can be used for a complex assessment of ripening stage. This was a reason for determination of relationship between both parameters. We found negative correlation between hardness and hue (hardness and hue on the cut coefficient -0.8447 ; hardness and hue on the surface coefficient -0.7459). On the other hand, the correlation between hardness and lightness was positive (hardness and lightness on the cut coefficient 0.6261; hardness and lightness on the surface coefficient 0.8606).

These characterize the relation between disappearing of unripe white curd with softening of the cheese mass. Ripening processes take place in the cheese and the hardness is decreasing and in the same time the appearance of the cheese is changing as well. There is a connection with the smear culture development which accelerates the proteolysis of milk proteins and produces colourful pigments. The changes are continuous with apparent trend and from the characteristics it is possible to estimate the degree of ripeness of the cheese and manage the quality system in the dairy plant and assure the quality of the product for the consumers.

4. Conclusions

Considering the fact that the appearance and hardness of the smear cheeses are the only characteristics which can be assessed by the customers direct during shopping, it is important to study these properties including exact instrumental measurements. From our results we can conclude that desired properties were achieved only after 38 days of ripening at the end of the ripening time, respectively on the expiry date. In this phase the cheese has typical colour on the surface and in the core of the cheese. White curd centre entirely disappears what conduce to significant decrease in lightness and to unification of yellow hue on the cheese surface and in the core. Ripening leads also to elastic, soft texture which can be recognized also by the squeezing between fingers (analogically to the instrumental test). The results of the

instrumental measurements are accurate, convenient for the statistical analysis and objective. It can be important for the quality control and monitoring of the ripening process in the dairy and it can predict how consumers evaluate the product.

5. References

- [1]Bockelmann, W. 2002. Development of defined surface starter cultures for the ripening of smear cheeses. In: *International Dairy Journal*, Vol. 12, 2002, s. 123–131.
- [2]Bockelmann, W. – Willems, K.P. – Neve, H. – Heller, K.H. 2005. Cultures for the ripening of smear cheeses. In: *International Dairy Journal*, Vol. 15, 2005, s. 719–732.
- [3]Cwíková, O. – Nedomová, Š. 2007. Vliv délky zrání olomouckých tvarůžků na jejich senzoričké a texturní vlastnosti. In: *Acta Universitatis agriculturae et silviculturae Mendeliana Brunensis*, Vol. 5, No. 55, 2007, s. 45–50.
- [4]Dufossé, L. – Galaup, P. – Carlet E. – Flamin, C. – Valla, A. 2005. Spectrocolorimetry in the CIE L*a*b* color space as useful tool for monitoring the ripening process and the quality of PDO red-smear soft cheeses. In: *Food Research International*, Vol. 38, 2005, s. 919–924.
- [5]Galaup, P. – Gaurtier, A. – Piriou, Y. – Villeblanche, A. – Valla, A. – Duffose, L. 2007. First pigment fingerprints from rind of French PDO red-smear ripened soft cheeses Epoisses. In: *Innovative Food Science and Emerging Technologies*, Vol. 38, 2007, 373–378.
- [6]Rao, R. – Toutenburg, H. – Shalabh – Heumann, C. 2008. *Linear Models and Generalizations, Least Squares and Alternatives*, 3rd edition, Springer, 2008. 572 s. ISBN: 3-642-09353-1
- [7]Sousa, M.J. – Ardo, Y. – McSweeney, P.L.H. 2001. Advances in the study of proteolysis during cheese ripening. In: *International Dairy Journal*, Vol. 11, No. 55, 2001, s. 327–345.
- [8]Standarová, E. – Vorlová, L. – Kordiovská, P. – Janštová, B. – Dračková, M. – Borkovcová, I. 2010. Biogenic Amine Production in Olomouc Curd Cheese (Olomoucké tvarůžky) at Various Storage Conditions. In: *Acta Veterinaria Brno*, Vol. 79, 2010, s. 147–156.

This work was financially supported by a grant SP2A00071/2101 from Internal Grant Agency at Mendel University in Brno.

Adresa autorov:

David Hampel, Mgr. Ph.D.
Mendel university in Brno
ÚSO PEF
Zemědělská 1
613 00 Brno
Czech Republic
david.hampel.uso@mendelu.cz

Šárka Hanzelková, Ing.
National Institute of Public
Health in Prague
Center for Health, Nutrition and
Food
Palackého 3a
612 42 Brno
Czech Republic
sarka.hanzelkova@gmail.com

Aneta Bubeníčková, Ing.
Mendel university in Brno
ÚTP AF
Zemědělská 1
613 00 Brno
Czech Republic
xbubeni0@mendelu.cz

Vývoj tržieb v potravinárskom priemysle SR v období 2000 až 2010 Food Industry turnover development in the period from 2000 to 2010

Jozef Chajdiak, Kateřina Věntusová

Abstract: This paper analyses the development of food industry in Slovakia in the period from 2000 to 2010 by volume turnover for the achievements and goods. In this analysis were used line and bar charts, moving annual averages, device: Add trend joining line, subsystem Contingent table from Excel. For the specification of the seasonal components were used indexes and deviations from the trend component by Moving annual averages, average values and were presented in a bar and line graphs.

Abstrakt: V príspevku sa analyzuje vývoj v potravinárskom priemysle SR v období 2000 až 2010 podľa objemu tržieb za vlastné výkony a tovar. K analýze boli použité spojnicové a stĺpcové grafy, kľzavé ročné priemery, nástroj Pridať trendovú spojnicu, podsystém Kontingenčná tabuľka z Excelu. K špecifikácii sezónnej zložky sa použili indexy a odchýlky od trendovej zložky vyjadrenej kľzavými ročnými priemermi, z nich vypočítané priemerné hodnoty a prezentované stĺpcovými a spojnicovými grafmi.

Key words: turnover, trend, seasonality, food industry

Kľúčové slová: tržby, trend, sezonnosť, potravinársky priemysel

JEL classification: C10, C19

Úvod

Ľudstvo v súčasnom historickom období prežíva dynamické zmeny. Počet obyvateľov sveta prekročil šesť miliárd, my sme sa prepracovali na úroveň 5,4 milióna obyvateľov. Úmerne k tomu rastie aj objem výroby potravín, nápojov a tabaku slúžiacich k uspokojeniu základných potrieb človeka. Vývojové zmeny v priemysle ovplyvňujú tak požiadavky vyplývajúce z uspokojovania základných potrieb človeka a ľudstva (dopyt), ako aj objem viazaných zdrojov v priemysle a efektívnosť ich spotreby. Analýza vývoja objemu tržieb za vlastné výkony a tovar za divízie 10 až 12 podľa Štatistickej klasifikácie ekonomických činností, Rev. 2 (2008), SK NACE Rev.2 – Výroba potravín, nápojov a tabaku, aj v porovnaní s údajmi za priemysel spolu tak predstavuje potrebnú súčasť analytických prác vo všeobecnosti.

1. Východiskové údaje

Štatistický úrad SR na svojej webovej stránke v časti Ukazovatele ekonomického vývoja SR, v časti Priemysel, v časti Časové rady vybraných ukazovateľov podľa SK NACE Rev. 2 (index priemyselnej produkcie, tržby, priemerný počet zamestnaných osôb, priemerná nominálna mesačná mzda, nové objednávky, odpracované hodiny, cenové indexy) prezentuje hodnoty uvedených ukazovateľov v podobe mesačných časových radov od januára 2000 po ostatné aktuálne obdobie. Údaje sú za priemysel spolu a väčšinu priemyselných divízií podľa SK NACE Rev.2. Za divízie 10 – 12 sme použili údaje o tržbách.

K vyjadreniu kumulatívnej veľkosti inflácie boli použité podklady z autorovho EIS Slovensko pričom východiskové podkladové údaje predstavujú údaje o mesačnej inflácii k rovnakému obdobiu predchádzajúceho roka.

Tržby za vlastné výkony a tovar autori vypočítali ako súčet publikovaných tuzemských a zahraničných tržieb.

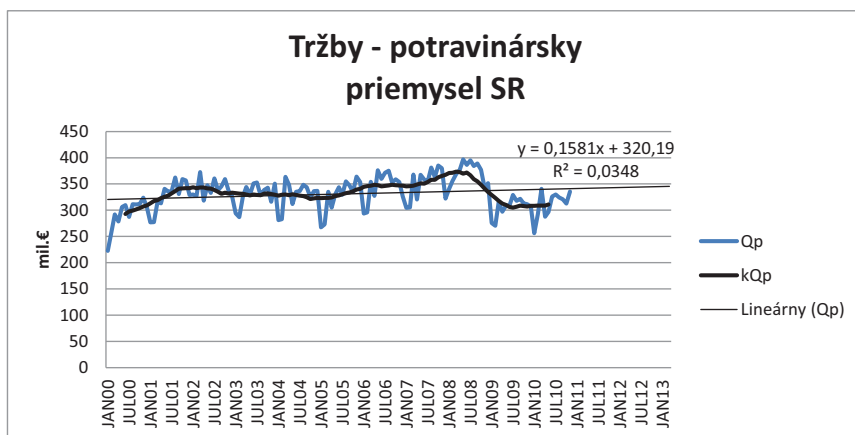
Tržby v cenách januára roku 2000 (očistené o vplyv inflácie) autori odhadli pre i-ty mesiac ako súčin nominálnych tržieb za i-ty mesiac krát kumulatívny index (inflácie) za január 2000 delene kumulatívnym index (inflácie) za i-ty mesiac.

Prehľad označení jednotlivých ukazovateľov je nasledujúci:

- Qp - Tržby za vlastné výkony a tovar
 kINF - Kumulatívna inflácia (k jan1989)
 $iQ_t = Q_t \cdot kINF_{jan2000} / kINF_t$, Tržby v cenách jan2000

2. Analýza vývoja objemu tržieb

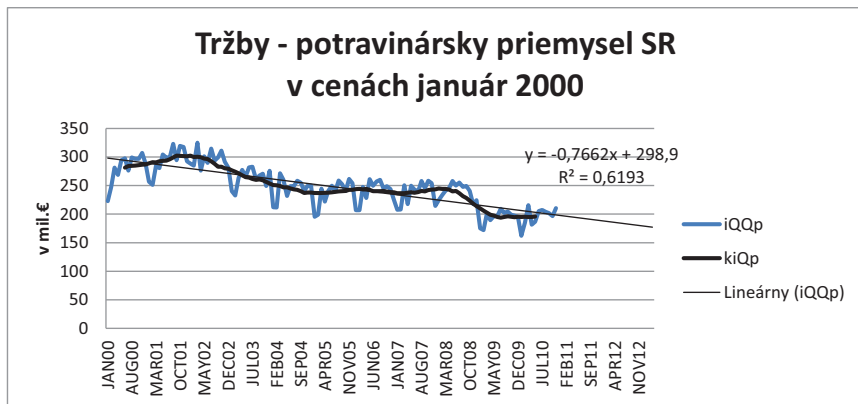
Na obr.1 je znázornený vývoj celkového objemu tržieb za vlastné výkony a tovar za divízie 10-12 SK NACE výroba potravín, nápojov a výrobkov z tabaku (potravinársky priemysel). Do apríla 2008 môžeme vidieť veľmi mierny rast objemu tržieb, nasleduje pokles v dôsledku hospodárskej kríz a od leta 2009 symbolický náznak rastu. Regresný koeficient 0,1581 modeluje prírastok objemu tržieb o 0,158 mil. € mesačne za predpokladu platnosti modelu regresnej priamky.



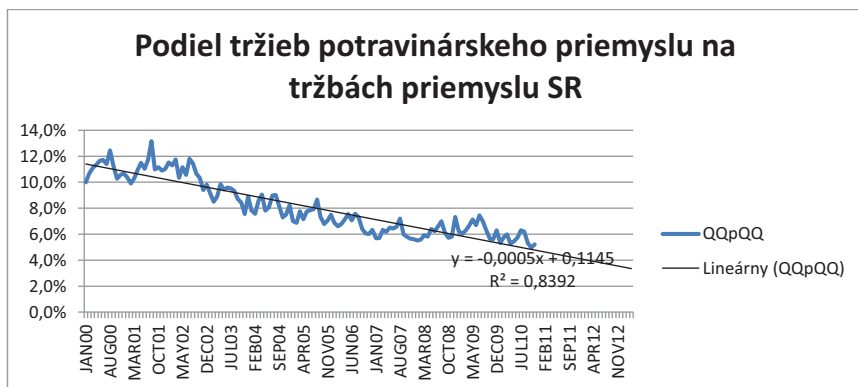
Obr.1

Na obr.2 je znázornený vývoj celkového objemu tržieb za vlastné výkony a tovar za divízie 10-12 SK NACE výroba potravín, nápojov a výrobkov z tabaku (potravinársky priemysel) v cenách januára 2000. Do marca 2002 môžeme vidieť postupný pokles objemu tržieb, nasleduje pokles v dôsledku hospodárskej kríz a niekoľko mesiacov na konci obdobia symbolický náznak rastu. Regresný koeficient -0,7662 modeluje pokles objemu tržieb o 0,766 mil. € mesačne za predpokladu platnosti modelu regresnej priamky.

Na obr.3 je znázornený vývoj podielu tržieb potravinárskeho priemyslu na tržbách priemyslu spolu. Podiel sa dlhodobo znižuje, podľa zobrazeného modelu regresnej priamky o 0,05 % mesačne.



Obr.2



Obr.3

3. Analýza sezónnosti tržieb v potravinárskom priemysle

Dôležitým aspektom v analýze vývoja hodnôt zvoleného ekonomického ukazovateľa je analýza sezónnosti. K posúdeniu sezónnej zložky sa používajú individuálne sezónne indexy (použiteľné len v prípade, že všetky hodnoty v časovom rade sú kladné) alebo sezónne diferencie (použiteľné v prípade ľubovoľných hodnôt). V dekompozícii časového radu trendovú zložku vyjadríme kľzavým ročným priemerom hodnôt časového radu a môžeme vypočítať individuálne sezónne indexy:

$$Y_t = Tr_t * S_t * E_t$$

Keďže stredná hodnota náhodnej chyby $E[E_t]=1$, individuálny sezónny index vypočítame si

$$s_i = S_i = \frac{Y_i}{Tr_i} = \frac{Y_i}{kY_i}$$

kde je Y_t – hodnota časového radu Y v časovom momente t ,

Tr_t – trendová zložka v čase t ,

S_t – sezónna zložka v čase t ,

E_t – náhodná zložka v čase t ,

kY_t – odhad hodnoty trendovej zložky Tr_t kľzavým ročným priemerom kY_t časového radu Y v čase t .

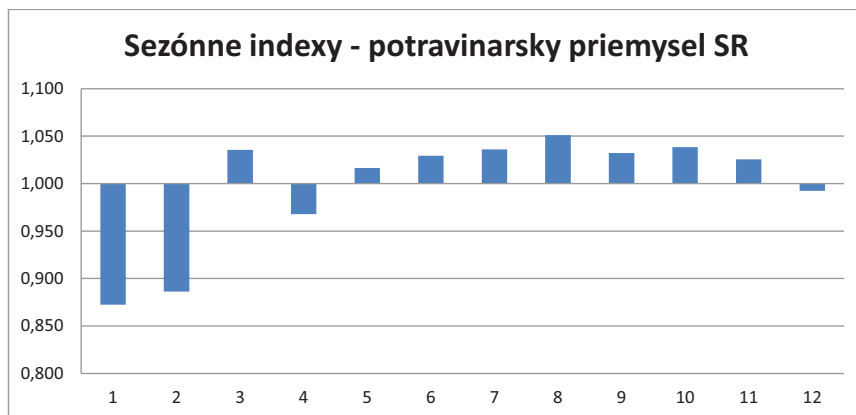
Individuálne sezónne indexy preusporiadame z tvaru časového radu do tvaru tabuľky, v ktorej stĺpce predstavujú roky a riadky predstavujú mesiace. Hodnoty riadkov agregujeme do sezónneho indexu vyjadrujúceho sezónny vplyv príslušného mesiaca. K agregácii použijeme funkciu geometrického priemeru. K inému vyjadreniu hodnôt sezónnych indexov použijeme mediány z príslušných mesačných hodnôt a k vyjadreniu variability hodnôt individuálnych indexov v príslušných mesiacoch použijeme minimálnu a maximálnu hodnotu. Vypočítané hodnoty individuálnych indexov a agregovaných hodnôt za potravinársky priemysel SR v tab.1.

Tab.1 Sezónne indexy v potravinárskom priemysle SR

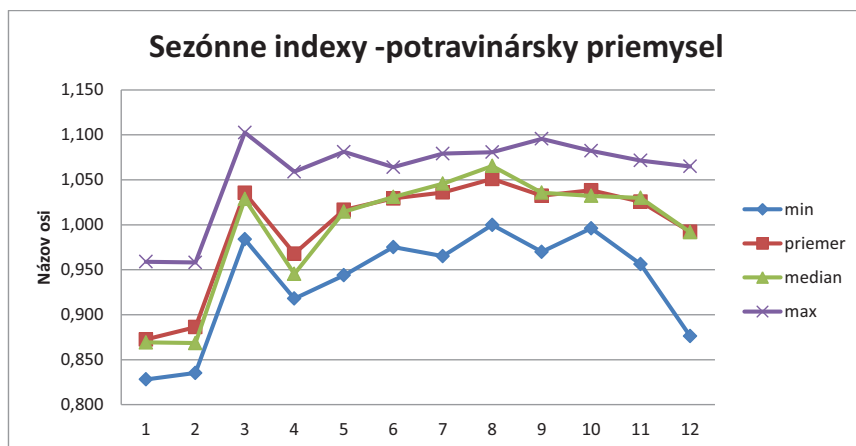
Me- siac	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	min	prie- mer	me- dián	max
1		0,884	0,959	0,885	0,858	0,828	0,853	0,881	0,919	0,837	0,831	0,828	0,872	0,869	0,959
2		0,873	0,956	0,864	0,858	0,844	0,855	0,883	0,958	0,835	0,947	0,835	0,886	0,868	0,958
3		0,991	1,087	0,984	1,101	1,035	1,022	1,061	0,987	0,993	1,103	0,984	1,035	1,029	1,103
4		0,969	0,928	1,043	1,059	0,940	0,939	0,918	1,011	0,951	0,932	0,918	0,968	0,945	1,059
5		1,046	1,018	1,005	0,944	1,011	1,081	1,047	1,074	0,995	0,954	0,944	1,017	1,015	1,081
6	1,056	1,023	0,975	1,064	1,019	1,048	1,041	1,022	1,038	1,010		0,975	1,029	1,031	1,064
7	0,965	1,012	1,066	1,072	1,026	0,997	1,073	1,000	1,078	1,079		0,965	1,036	1,046	1,079
8	1,041	1,077	1,010	1,000	1,066	1,069	1,081	1,065	1,069	1,037		1,000	1,051	1,065	1,081
9	1,032	0,970	1,044	1,020	1,059	1,040	1,013	1,014	1,096	1,042		0,970	1,032	1,036	1,096
10	1,026	1,052	1,078	1,032	1,011	0,996	1,033	1,061	1,082	1,017		0,996	1,038	1,032	1,082
11	1,055	1,042	1,015	0,956	1,043	1,071	1,020	1,040	1,003	1,015		0,956	1,026	1,030	1,071
12	0,987	0,963	0,981	1,065	1,041	1,038	0,941	0,876	1,050	0,996		0,876	0,992	0,992	1,065
Spo- lu	1,023	0,992	1,010	0,999	1,007	0,993	0,996	0,989	1,030	0,984	0,953	0,953	0,998	0,996	1,030

Analýzu množstva čísel s hodnotami okolo 1 uľahčuje ich grafická prezentácia. Jednoduchú prezentáciu obsahujú obr.4 a obr.6 so stĺpcovými grafmi vyjadrujúcimi sezónne indexy. Dôležitým aspektom prezentácie je, že os x prechádza cez počiatok nie v bode nula ale v bode 1 na osi y . Tak body nad osou x sú sezónne nadpriemerné a body pod osou x sú sezónne podpriemerné. Druhý typ grafov predstavuje spojnicový graf, pričom na osi x máme príslušné sezónne obdobia a spojnicami sú spojené hodnoty minim, priemerov, mediánov a maxim.

Pri analýze grafu na obr.4 vidíme dva výrazne podpriemerné mesiace 1 – január a 2 – február. Podpriemerný je aj 4 – apríl a 12 – december. Mesiac 5 – máj až 11 -november predstavuje postupnosť nadpriemerných mesiacov s vrcholom v 8 – august. Obr.5 graficky prezentuje stabilitu sezónnych indexov na obr.4. Čím je variačné rozpätie medzi minimom a maximom väčšie, tým sú individuálne sezónne indexy variabilnejšie a tým pre riadenie vývoja časového radu menej použiteľné. Vidíme, že spojnica minim a maxim kopiruje spojnicu priemerov resp. mediánov. V potravinárskom priemysle najmenej spoľahlivý z hľadiska interpretácie je mesiac 12 - december.

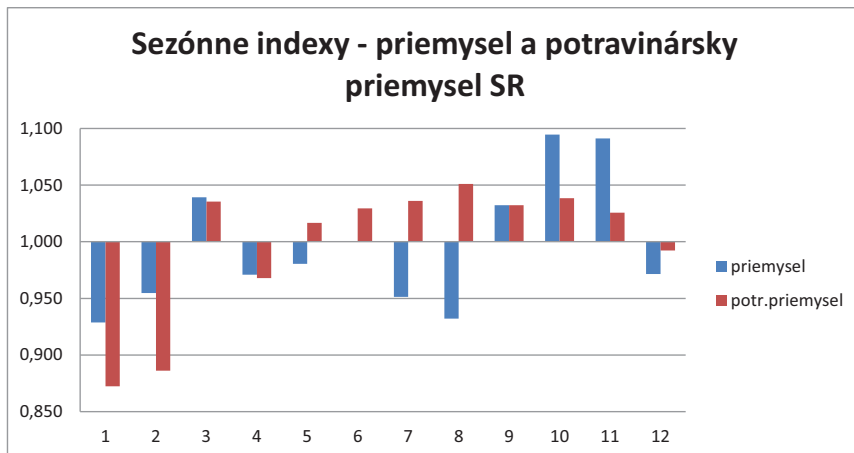


Obr.4



Obr.5

Na obr. 6 sú spoločne prezentované sezónne indexy v priemysle a v potravinárskom priemysle. Oboch súboroch môžeme konštatovať zhodnú podpriemernosť v mesiacoch december, január, február a apríl a nadpriemernosť v mesiacoch september, október a november. V letných mesiacoch v priemysle vrcholia dovolenky a v potravinárskom priemysle zberová sezóna.



Obr.6

Záver

Z vývoja ukazovateľa tržby za vlastné výkony a tovar môžeme konštatovať skoro stagnujúci vývoj v potravinárskom priemysle. Sezónny vývoj v zimných mesiacoch je podobný v priemysle a potravinárskom priemysle, v letných mesiacoch je tento charakter sezónnosti naopak – potravinársky priemysel realizuje nadpriemerne tržby, kým priemysel spolu podpriemerne.

Literatúra

[1] Chajdiak, J. 2003. *Štatistika jednoducho. 3. Vydanie*. Bratislava: Statis, 2010. 194 s. ISBN 80-85659-60-3.

[2] Chajdiak, J. 2009. *Štatistika v Exceli 2007*. Bratislava: Statis, 2009. ISBN 80-85659-49-8.

Vypracované v rámci riešenia úlohy VEGA č.1/0536/10 „Inovácia ako strategický základ zvyšovania konkurenčnej schopnosti SR“.

Adresa autorov:

Jozef Chajdiak, Doc.Ing.CSc.
Ústav manažmentu STU
Bratislava
jozef.chajdiak@stuba.sk

Mgr. Kateřina Věntusová
Výskumný ústav potravinársky
Bratislava
katerina.ventusova@vup.sk

Podobnosť krajín EÚ z hľadiska vnímania korupcie Similarity of EU countries in terms of corruption perception

Jaroslav Ivor, Beáta Stehlíková, Anna Tirpáková

Abstract: The article analyzed the similarity of the European Union on the basis of fuzzy representation of the CPI. To create clusters, the European Union in terms of similarity of the corruption expressed of the Corruption Perceptions Index (CPI) we used cluster analysis, where the CPI data are modeled as triangular fuzzy number.

Key words: corruption, Corruption Perception Index, triangular fuzzy number, cluster analysis

Kľúčové slová: korupcia, Index vnímania korupcie, trojuholníkové fuzzy číslo, zhluková

JEL classification: C10

1. Úvod

EÚ urobila z boja proti korupcii jednu zo svojich najvyšších politických priorit – ako vo vzťahu k iným krajinám, tak aj smerom dovnútra [9]. *Svetová banka* identifikuje korupcie ako „najväčšiu brzdú ekonomického a sociálneho rozvoja“. Podľa nej podkopáva rozvoj narúšaním vlády zákona a oslabovaním inštitucionálnych základov, na ktorých stojí hospodársky rast. Škodlivé dopady korupcie sú mimoriadne negatívne voči chudobným, ktorých ekonomický pokles poškodzuje najviac. Chudobní sú najviac závislí na poskytovaní verejných služieb a sú najmenej schopní niesť dodatočné náklady z poskytovania úplatkov, podvodov a zneužívania ekonomických privilégií.

Pod pojmom korupcia máme dnes na mysli predovšetkým porušovanie právnych a mravných noriem, ktorého cieľom je zvýhodnenie jednotlivca či skupiny jednotlivcov oproti ostatným občanom [2]

V Občianskoprávnom dohovore o korupcii Rady Európy, sa v Článku 2 konštatuje, že korupcia znamená žiadanie, ponúkanie, poskytovanie alebo prijímanie, priamo alebo nepriamo, úplatku alebo inej nenáležitej výhody alebo ich príslubu, ktoré deformuje riadne vykonávanie povinností alebo správanie požadované od príjemcu úplatku, nenáležitej výhody alebo ich príslubu. V dôvodovej správe [8] k Dohovoru 370/2003 Z.z. sa konštatuje: "že účelom tejto definície bolo zabezpečiť to, aby z práce GMC nebolo nič vylúčené. Zjavne, táto definícia by sa nevyhnutne nemusela hodiť k právnej definícii korupcie vo väčšine členských štátov, a to najmä k trestnoprávnej definícii, ale jej výhodou bolo, že dopredu neobmedzovala diskusiu príliš úzkymi hranicami." Korupčné konanie z trestnoprávneho hľadiska slovenského práva vymedzuje Trestný zákon 300/2005 Z.z. v treťom diele ôsmej hlavy, ktorá nesie názov Trestné činy proti poriadku vo veciach verejných. Pod pojem korupcia pritom spadajú tri skutkové podstaty trestného činu, a to Prijímanie úplatku alebo inej nenáležitej výhody, dopĺňanie a nepriama korupcia.

V Občianskoprávnom dohovore o korupcii Rady Európy, sa v Článku 2 konštatuje, že korupcia znamená žiadanie, ponúkanie, poskytovanie alebo prijímanie, priamo alebo nepriamo, úplatku alebo inej nenáležitej výhody alebo ich príslubu, ktoré deformuje riadne vykonávanie povinností alebo správanie požadované od príjemcu úplatku, nenáležitej výhody alebo ich príslubu. V dôvodovej správe [8] k Dohovoru 370/2003 Z.z. sa konštatuje, " že účelom tejto definície bolo zabezpečiť to, aby z práce GMC nebolo nič vylúčené. Zjavne, táto definícia by sa nevyhnutne nemusela hodiť k právnej definícii korupcie vo väčšine členských

štátov, a to najmä k trestnoprávnej definícii, ale jej výhodou bolo, že dopredu neobmedzovala diskusiu príliš úzkymi hranicami."

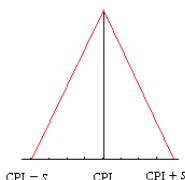
Korupcia je vážny celospoločenský problém, ktorého riešenie je dôležité a aktuálne, konštatujú, Ďurajková a i. [1]. Na čoraz viac globalizovanom trhu, možnosti a rozsah korupcie, rastú. Bohatstvo veľkých spoločností poskytuje zdroje pre korupciu. Liberalizácia medzinárodných finančných trhov umožňuje ukrytie výnosov z korupcie. Významné korupčné transakcie zostávajú skryté pred zrakmi verejnosti, ale ekonomické dôsledky, spoločenské a politické dôsledky korupcie dopadajú na celú spoločnosť.

2. Materiál a metódy

Transparency International zostavuje Index vnímania korupcie (CPI) každoročne od roku 1995. Nadobúda hodnoty od nula (najnižšia hodnota vnímania korupcie) po 10 (najvyššia hodnota vnímania korupcie). Od roku 2001 uvádza aj štandardnú odchýlku tohto ukazovateľa, resp. 90 percentný interval spoľahlivosti - Index The vast majority of the 180 countries included in the 2009 [7]. V tabuľke 1 sú uvedené hodnoty Indexu vnímania korupcie (CPI) s príslušným intervalom spoľahlivosti štátov EÚ.

Meranie je chápané vo všeobecnej rovine ako spôsob priradenia hodnoty charakteristik jednotlivým objektom. Merateľ v *F škále* znamená určovanie stupňa príslušnosti jednotlivých objektov. *Indexy* vznikajú špeciálnou konštrukciou z indikátorov charakteristik. Proces konštrukcie indexu je často jediným prostriedkom *sformovania nového pojmu na empirickej úrovni poznatkov* alebo záměny nepresného teoretického pojmu presným. Konštrukciu indexu môžeme vnímať aj ako spôsob vyjadrenia latentného znaku, ktorý sa nedá bezprostredne merať. Údaje CPI sme v [5] modelovali ako trojuholníkové fuzzy číslo. Trojuholníkové fuzzy číslo A je definované na intervale $\langle a_1, a_2 \rangle$ a jeho funkcia príslušnosti má tvar

$$\mu_A(x) = \begin{cases} \frac{x - a_1}{a_M - a_1}, & a_1 \leq x \leq a_M \\ \frac{x - a_2}{a_M - a_2}, & a_M \leq x \leq a_2 \\ 0, & \text{inak} \end{cases}$$



Zdroj: Vlastné zobrazenie

Obrázok 1. CPI ako trojuholníkové fuzzy číslo

Vzdialenosť dvoch fuzzy čísel $A = (a_1, a_M, a_2)$ a $B = (b_1, b_M, b_2)$ je definovaná pomocou euklidovskej vzdialenosti.

Pre vytvorenie zhlukov štátov Európskej únie z hľadiska podobnosti vnímania korupcie vyjadrenej pomocou CPI sme použili zhlukovú analýzu [3]. Zhluková analýza zahŕňa mnoho matematických metód, ktoré môžu byť použité pre identifikáciu skupín objektov, ktoré sú si podobné. Wardova metóda, ktorú sme použili, hierarchického zhlukovania, spája

pozorovania do skupín tak, aby medzizhlukový rozptyl bol čo najnižší. Táto metóda sa zreteľne odlišuje od všetkých ostatných pretože na určenie vzdialenosti medzi zhlukmi využíva prístup analýzy rozptylu. Pomocou tejto metódy sa zhluky vytvárajú tak, aby sa vnútrozhlukový súčet štvorcov minimalizoval. Celý proces zhlukovania možno názorne sumarizovať do tzv. dendrogramu, ktorý predstavuje grafické znázornenie podobnostného stromu.

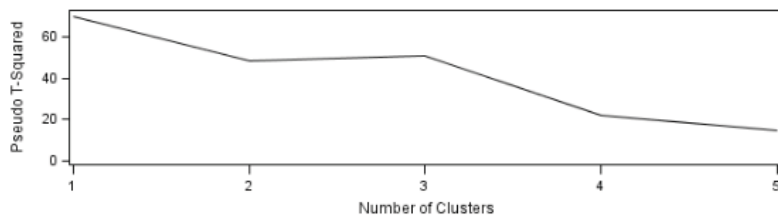
3. Výsledky a diskusia

Obsah pojmu korupcie nie je ustálený a jednoznačne vymedzený. Preto aj s interpretáciou CPI treba byť opatrný vzhľadom k tomu, zhromaždené údaje sa týkajú vnímania a nie skutočných javov. Korupčné javy je potrebné posudzovať v kontexte kultúrnych a sociálnych reálií konkrétnych štátov, etnických či národných spoločenstiev, vo vzťahu k tradícii, verejnej mienke, právnomu systému a podobne. To, čo je v jednom kontexte kvalifikované ako korupčné a morálne odsudzované ako aj právne sankcionované, v iných kontextoch je bežnou legálnou praxou. Pozorovaná vysoká miera korupcie pomocou CPI môže odrážať vysoký štandard etiky ale tiež vysokú mieru reálneho korupčného správania. Vnímanie CPI bez širšieho kontextu o danej krajine tak môže byť zavádzajúci ukazovateľ o skutočnej úrovne korupcie.

Tabuľka 1. Index vnímania korupcie (CPI) ako fuzzy číslo

Štát	CPI 2009 a_M	90% Interval spoľahlivosti	
		Dolná hranica a_1	Horná hranica a_2
Bulharsko	3,8	3,2	4,5
Grécko	3,8	3,2	4,3
Rumunsko	3,8	3,2	4,3
Taliansko	4,3	3,8	4,9
Lotyšsko	4,5	4,1	4,9
Slovensko	4,5	4,1	4,9
Česká republika	4,9	4,3	5,6
Litva	4,9	4,4	5,4
Poľsko	5	4,5	5,5
Maďarsko	5,1	4,6	5,7
Malta	5,2	4	6,2
Portugalsko	5,8	5,5	6,2
Španielsko	6,1	5,5	6,6
Cyprus	6,6	6,1	7,1
Estónsko	6,6	6,1	6,9
Slovinsko	6,6	6,3	6,9
Francúzsko	6,9	6,5	7,3
Belgicko	7,1	6,9	7,3
Spojené kráľovstvo	7,7	7,3	8,2
Rakúsko	7,9	7,4	8,3
Nemecko	8	7,7	8,3
Írsko	8	7,8	8,4
Luxembursko	8,2	7,6	8,8
Fínsko	8,9	8,4	9,4
Holandsko	8,9	8,7	9
Švédsko	9,2	9	9,3
Dánsko	9,3	9,1	9,5

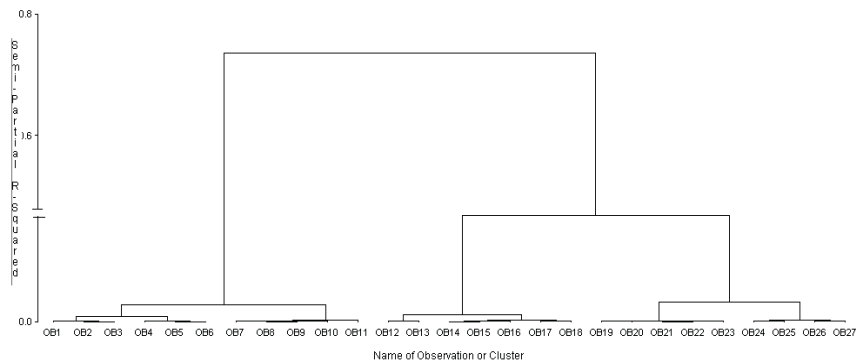
Zdroj: [7]



Zdroj: Vlastné výpočty a zobrazenie

Obrázok 2. Určenie optimálneho počtu zhlukov

Na základe hodnôt Pseudo T-squared je počet optimálnych zhlukov dva. Dendrogram výsledku zhlukovania je na nasledujúcom obrázku.



Zdroj: Vlastné výpočty a zobrazenie

Obrázok 3. Zhluky štátov Európskej únie z hľadiska vnímania korupcie vyjadrenej pomocou CPI.

Do prvého zhluku patria štáty Bulharsko, Grécko, Rumunsko, Taliansko, Lotyšsko, Slovensko, Česká republika, Litva, Poľsko, Maďarsko, Malta. S výnimkou Malty ide o postkomunistické štáty. Druhý zhluk tvoria dve skupiny štátov - Portugalsko, Španielsko, Cyprus, Estónsko, Slovinsko, Francúzsko a Belgicko a skupina štátov Spojené kráľovstvo, Rakúsko, Nemecko, Írsko, Luxembursko, Fínsko, Holandsko, Švédsko a Dánsko.

4. Záver

Príčiny korupcie súvisia s vysokým prospechom z korupcie a nízkym rizikom spojeným s korupčnou transakciou. Čím je možnosť prospechu vyššia, tým je motivácia konať korupčne vyššia. Zároveň, čím menej hrozí, že korupcia bude odhalená a potrestaná, tým viac odvahy existuje na korupčné konanie [1].

Dosiahnutý výsledok potvrdzuje výsledok prezentovaný v na infowebe, že skutočne existuje určitý vzťah medzi celkovou ekonomickou situáciou krajiny a rozšírením korupcie. V tomto vzťahu môžeme hospodársku situáciu nahradiť aj demokraciou a všimnúť si, že aj tu existuje nepriama väzba na korupciu. Korupcia je totiž minimálne rozšírená v krajinách so stabilne fungujúcou demokraciou.

5. Literatúra

- [1] ĎURAJKOVÁ, D.- ZEMANOVIČOVÁ, D.- ZUBERSKÁ, M. 2004. Celospoločenský problém – korupcia II. Transparency International Slovensko, Bratislava 2004
- [2] HERZOGOVÁ, Z. 2003. Policejní Etika. 1 vyd. Praha: Vydané vlastným nákladom, 2003. s. 167-168. ISBN 80-238-9984-8
- [3] IRENE, M. M.: Clustering Methods and Algorithms. Dostupné na internete: <http://www.cse.iitb.ac.in/dbms/Data/Courses/CS632/1999/clustering/dbms.html>
- [4] SCOTT, J.C. 1969. *The Analysis of Corruption in Developing Nations*. Comparative Studies in Society and History, 11, pp 315-341
- [5] STEHLÍKOVÁ, B.- TIRPÁKOVÁ, A. 2010.: Vnímanie korupcie ako fuzzy číslo = Corruption perception as a fuzzy number. In: Ekonomická kríza: nové podnety pre ekonomickú teóriu a prax, Zborník z medzinárodnej vedeckej konferencie: Economic crisis: new incentives for the economic theory and practice, The proceedings of the international scientific conference. – Bratislava, Eurokódex, 2010. - ISBN 9788089453016. - s. 248-257.
- [6] WRAITH, R.- SIMPKINS, E. 1963. *Corruption in Developing Countries*. London: George Allen & Unwin Ltd., 211 p. 30s. net.
- [7] Corruption Perceptions Index 2009
http://www.transparency.org/policy_research/surveys_indices/cpi/2009/cpi_2009_table
- [8] Občianskoprávny dohovor o korupcii. Štrasburg, 4.11.1999, Rada Európy, V SR Dohovor 370/2003 Z.z., Dôvodová správa k Dohovoru 370/2003 Z.z
http://www.justice.gov.sk/dwn/r6/05/dbpk_obc_ds_rtf
- [9] Korupcia
http://www.euractiv.sk/verejna-sprava/zoznam_liniek/korupcia
- [10] Korupcia ako determinant regionálneho rozvoja.
<http://www.infoweby.sk/spolocnost/regionalny-rozvoj/61-korupcia-ako-determinant-regionalneho-rozvoja>

Adresa autorov:

Jaroslav Ivor, Prof. JUDr., DrSc.
Fakulta práva BVŠP
Tomášiková 20
851 05 Bratislava
dekan.fp@uninova.sk

Beáta Stehlíková, prof. RNDr., CSc.
Fakulta ekonomie a podnikania BVŠP
Tematínska 10
851 05 Bratislava
stehlikovab@gmail.com

Anna Tirpáková, prof. RNDr., CSc.
Katedra Matematiky UKF
Tr. A. Hlinku 1
949 01 Nitra
atirpakova@ukf.sk

Príspevok je financovaný z projektu FCVV FPV UKF Nitra.

Porovnanie miery nezamestnanosti v Košickom regióne a v Slovenskej republike

The comparison of the unemployment rate in Kosice region and Slovak republic

Gabriela Ižaríková, Denisa Olekšáková

Abstract: The aim of this paper is to compare the unemployment rate in Kosice region and the unemployment rate in Slovak republic. We use some statistical characteristics, for example arithmetic average, standard deviation and variation coefficient for the comparison of the unemployment rate. Pearson correlation coefficient shows that between the number of unemployed in Kosice region and the number of unemployed in Slovak republic is the linear dependence.

Key words: unemployment, unemployment rate, Kosice region

Kľúčové slová: nezamestnanosť, miera nezamestnanosti, Košický kraj

JEL classification: C10, C13, C19

1. Úvod

Daný príspevok je zameraný na pozíciu Košického kraja v regionálnej štruktúre Slovenska, konkrétne na nezamestnanosť. Poukazuje na to, či dochádza alebo nedochádza k prehĺbovaniu regionálnych rozdielov na Slovensku a ako sa v tomto kontexte rozvíja Košický kraj. Aplikované budú niektoré nástroje štatistiky, ako napr. aritmetický priemer, smerodajná odchýlka, variačný koeficient, korelačný koeficient.

Cieľom príspevku je prieskum nezamestnanosti v Košickom kraji v rámci Slovenska a taktiež analýza vnútorných diferenciácií v rámci kraja. Pod regionálnymi disparitami rozumieme rozdiely v stupni sociálno-ekonomického rozvoja regiónov, ktoré sú dôsledkom jeho nerovnomernosti. Pri tejto analýze je dôležitý výber vhodných ukazovateľov a výber vhodných štatistických nástrojov, umožňujúcich časopriestorové porovnávanie. Na meranie rozdielov možno použiť viacero štatistických nástrojov a mier (smerodajná odchýlka, variačný koeficient, Pearsonov korelačný koeficient, index korelácie, koeficienty rastu a poklesu, Theilov koeficient, Giniho index, MSE - Mean Squared Error, RMSE- Root Mean Squared Error). V tomto príspevku použijeme aritmetický priemer, smerodajnú odchýlku, variačný koeficient a korelačný koeficient.

Aritmetický priemer je najčastejšie používanou charakteristikou polohy, pri jeho výpočte sa využívajú všetky zistené hodnoty analyzovaného štatistického znaku. Aritmetický priemer je podiel súčtu všetkých hodnôt znaku a ich počtu a vyjadrujeme ho vzťahom

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (1)$$

Smerodajná (štandardná) odchýlka patrí medzi absolútne charakteristiky variability, dá sa interpretovať dvoma spôsobmi, ako priemerný rozdiel medzi hodnotami a priemerom pri ignorovaní znamienok, alebo ako priemerný rozdiel medzi každými dvoma hodnotami pri ignorovaní znamienok. Čím je väčšia variabilita hodnôt, tým sú väčšie aj odchýlky. Smerodajnú odchýlku charakterizuje vzťah

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2)$$

Variačný koeficient je relatívnu mierou disperzie odvodenou od štandardnej odchýlky (podiel štandardnej odchýlky a priemeru vyjadrený v percentách). Používa sa na porovnávanie variability medzi súbormi dát s odlišnými priermi a zapisujeme ho vzťahom

$$k = \frac{s}{\bar{x}}. \quad (3)$$

Pearsonov korelačný koeficient meria silu lineárnej závislosti dvoch číselných premenných, nadobúda hodnoty z intervalu $(-1,1)$. Podľa jeho znamienka sa posudzuje smer závislosti ($r_{xy} > 0$ - priama závislosť, $r_{xy} < 0$ - nepriama závislosť). Podľa jeho absolútnej hodnoty sa posudzuje sila závislosti obidvoch premenných ($r \leq 0,3$ nepreukázaná, $0,3 < r \leq 0,5$ mierna, $0,5 < r \leq 0,7$ vysoká, $0,7 < r \leq 0,9$ významná, $0,9 < r$ veľmi tesná). Pre korelačný koeficient platí

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{x\bar{y} - \bar{x} \cdot \bar{y}}{\sqrt{(x^2 - \bar{x}^2) \cdot (y^2 - \bar{y}^2)}}. \quad (4)$$

2. Košický kraj – charakteristika

Košický kraj je s rozlohou 6 755 km² štvrtý najväčší a zaberá 13,8% územia Slovenskej republiky. Podľa územno-správneho usporiadania sa člení na 11 okresov: Gelnica, Košice I, Košice II, Košice III, Košice IV, Košice-okolie, Michalovce, Rožňava, Sobrance, Spišská Nová Ves a Trebišov. V kraji je 440 obcí, z toho 17 má štatút mesta. Správnym, hospodárskym, politickým, školským a kultúrnym centrom kraja sú Košice, ktoré sú druhým najväčším mestom na Slovensku. Na území kraja celkom 2010 žilo celkom 780 000 obyvateľov, čo je 14,3% slovenského úhrnu. Týmto podielom je Košický kraj druhým najväčším na Slovensku. Kraj patrí k hustejšie osídleným regiónom, na 1 km² tu žije priemerne 115,2 obyvateľov, o 4,6 obyvateľov viac než je priemer. Priemerný vek obyvateľa kraja je 37,2 rokov, čo je o 1,3 roka menej oproti celoslovenskému priemeru.

V kraji sa nachádzajú energetické, rudné i nerudné suroviny. Z energetických sú to ropa a zemný plyn v okresoch Michalovce a Trebišov, zdroje geotermálnej energie sú v okrese Košice-okolie, z rudných surovín sú to strieborné rudy v okrese Rožňava a Spišská Nová Ves. V kraji sa taktiež nachádzajú rôzne druhy stavebného kameňa, tehliarske hliny, vápenec, kaolín, štrkopiesky a iné. Ťažiskovými ekonomickými odvetviami sú priemysel (hutnícky, strojársky, potravinársky, elektrotechnický, ťažobný, stavebných hmôt, palív a energetiky), stavebníctvo a poľnohospodárstvo. V kraji je rozvinutá aj sieť obchodov, služieb, finančných inštitúcií, zdravotníckych a sociálnych zariadení a vedeckých a kultúrnych ustanovizní.

3. Nezamestnanosť

Podľa všeobecne daných poučiek je nezamestnanosť sociálno-ekonomický jav vyjadrujúci neúplné využitie práceschopného obyvateľstva, ktoré má o prácu záujem (môže byť dobrovoľná, kedy je počet voľných pracovných miest väčší alebo rovnaký ako počet nezamestnaných ľudí, alebo nedobrovoľná).

Rozlišujú sa tri druhy nezamestnanosti:

- frikčná (materská dovolenka, presťahovanie sa),
- štruktúrna (neadekvátna kvalifikácia pre zamestnávateľa),
- cyklická (sezónny charakter výroby).

Nezamestnanosť je vlastne nerovnováha na trhu práce, ponuka práce je menšia ako dopyt po nej. Za nezamestnaného je v ekonomickej teórii považovaná osoba schopná pracovať, ktorá si však nemôže nájsť platené zamestnanie. Oficiálne (metodika Medzinárodnej organizácie práce "ILO") sa za nezamestnaného považuje osoba, ktorá:

- je staršia ako 15 rokov a nie je pracujúca,
- aktívne hľadá prácu,
- je schopná a ochotná nastúpiť do práce.

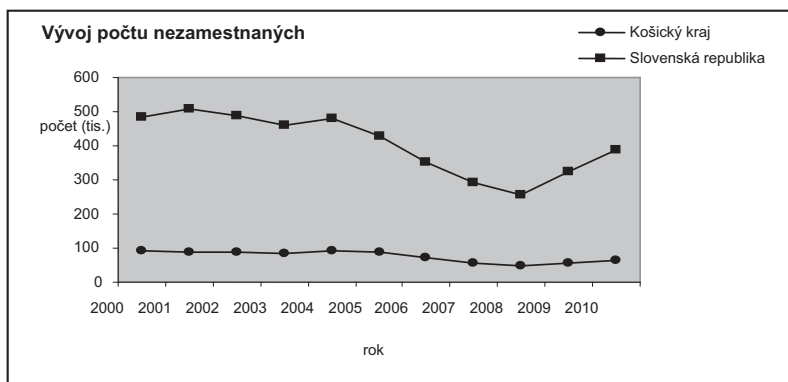
Tabuľka č. 1: Vývoj počtu nezamestnaných (tis. osôb) v krajoch SR v r. 2000-2010

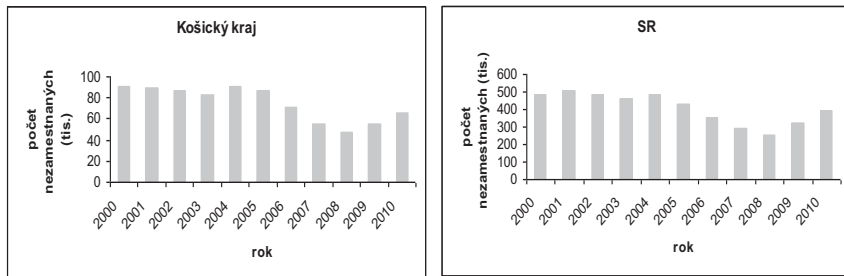
kraj/rok	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Bratislavský	24,4	28,2	28,7	22,8	27	17,1	14,4	14,1	12,4	16,3	20,9
Trnavský	44,7	51,5	45,7	37,1	36	30,1	25,4	18,7	18,3	27,5	36,6
Trenčiansky	44,2	39,8	33,1	27	25,4	23,9	21,2	16,7	13,9	21,4	30,7
Nitriansky	69,8	79,4	79,9	81,1	71	61,4	45,1	37,4	31,7	45,7	54,1
Žilinský	61,4	63,3	57,8	57,1	57,9	50,3	39,4	33,6	25,9	35,4	48,7
Banskobystrický	69,6	73,4	82,2	76,9	86,8	77,6	68,6	64,9	59,5	59,8	60,3
Prešovský	80,1	83,1	73,3	74,1	85,4	80,1	68	51,7	48,7	62,4	72
Košický	91,2	89,3	86,2	83,1	91,3	87	71,3	54,7	47,2	55,8	65,8
SR	485,4	508	486,9	459,2	480,8	427,5	353,4	291,8	257,6	324,3	389,1
aritmetický priemer	60,68	63,5	60,86	57,4	60,1	53,44	44,18	36,48	32,2	40,54	48,64

Zdroj: portal.statistics.sk

Počas sledovaného obdobia (2000 – 2010) počet nezamestnaných rastie alebo klesá, podľa jednotlivých krajov. Najnižší počet nezamestnaných je v Bratislavskom kraji, v roku 2008 ich bolo len 12,4 tisíc. V Košickom kraji bolo v rokoch 2000 až 2006 najviac nezamestnaných z celej Slovenskej republiky, v roku 2004 až 91,3 tisíc osôb. Príčiny a dôvody nezamestnanosti sú rôzne, napr. Košický kraj je druhý najväčší do počtu obyvateľov a aj husto obývaný. V roku 2007 toto prvenstvo prebral Banskobystrický kraj, ale v roku 2010 bolo najviac nezamestnaných opäť v Košickom kraji. V Košickom kraji je počet nezamestnaných vždy nad celoslovenským priemerom.

Hodnota Pearsonovho korelačného koeficientu 0,9625 svedčí o tom, že medzi počtom nezamestnaných v Košickom kraji a počtom nezamestnaných v Slovenskej republike je priama lineárna závislosť, číselná hodnota koeficientu naznačuje, že je veľmi tesná. Najväčšiu hodnotu Pearsonov korelačný koeficient nadobúda pre Žilinský kraj (0,9918) a najnižšiu hodnotu pre Banskobystrický kraj (0,7677). Hodnoty aritmetického priemeru, a Pearsonovho koeficientu pre dané roky sú vypočítané podľa vzťahov (1), (4).





Graf č. 1: Porovnanie vývoja počtu nezamestnaných v Košickom kraji a v SR

Z grafu č. 1 (spojnicový graf a histogram) je zrejmé, že krivky znázorňujúce počet nezamestnaných v Košickom kraji a počet nezamestnaných v SR majú takmer identický tvar aj trend len s odlišnosťou čísiel.

4. Miera nezamestnanosti

Ukazovateľ, ktorý hovorí o stave a vývoji nezamestnanosti sa nazýva miera nezamestnanosti. Miera nezamestnanosti je podiel nezamestnaných k ekonomicky aktívnym osobám (teda pracujúcim i nezamestnaným).

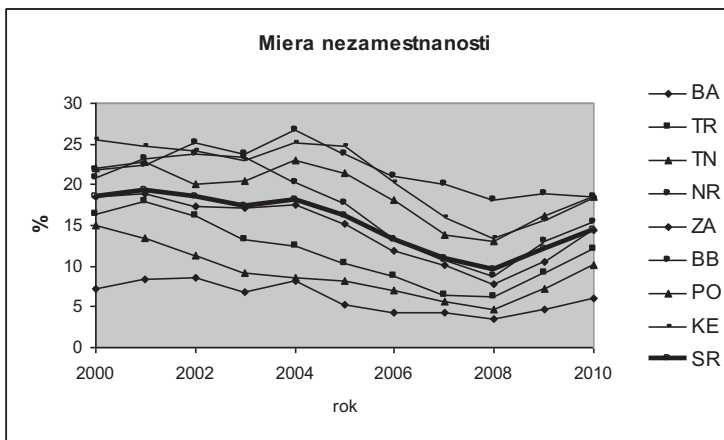
Tabuľka č. 2: Vývoj miery nezamestnanosti (v %) v krajoch SR v r. 2000-2010

kraj/rok	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Bratislavský	7,2	8,3	8,6	6,9	8,2	5,2	4,3	4,2	3,6	4,7	6,1
Trnavský	16,4	18	16,1	13,2	12,5	10,4	8,8	6,5	6,2	9,1	12
Trenčiansky	15	13,4	11,3	9,2	8,6	8,1	7,1	5,7	4,7	7,3	10,2
Nitriansky	20,8	23,1	23,8	23,4	20,3	17,8	13,2	10,7	8,8	13	15,4
Žilinský	18,5	18,9	17,3	17,2	17,5	15,2	11,8	10,1	7,7	10,6	14,5
Banskobystrický	21,9	22,4	25,2	23,8	26,6	23,8	21,1	20	18,2	18,8	18,6
Prešovský	22,1	22,7	20,1	20,4	22,9	21,5	18,1	13,8	13	16,2	18,6
Košický	25,6	24,8	24,1	23	25,2	24,7	20,3	15,9	13,5	15,5	18,3
SR	18,6	19,2	18,5	17,4	18,1	16,2	13,3	11	9,6	12,1	14,4
<i>aritmetický priemer</i>	18,44	19	18,31	17,14	17,73	15,84	13,09	10,86	9,463	11,9	14,21
<i>smerodajná odchýlka</i>	5,298	5,27	5,741	6,244	6,795	6,9	5,857	5,101	4,696	4,511	4,235
<i>varičný koeficient</i>	0,287	0,28	0,313	0,364	0,383	0,436	0,448	0,47	0,496	0,379	0,298

Zdroj: portal.statistics.sk

Miera nezamestnanosti na Slovensku v prvej časti sledovaného obdobia (2000-2004) stúpala, kulminovala v roku 2004. Od roku 2004 možno konštatovať pokles miery nezamestnanosti ale od roku 2009 opäť začína rásť. Tento vývoj je však regionálne diferencovaný, poukazuje na to hodnota variačného koeficientu, ktorý stúpol z hodnoty 0,28 v roku 2001 až na hodnotu 0,496 v roku 2008. Z grafu č. 2 je zreteľné, že najvyššia miera nezamestnanosti je v sledovanom období v Banskobystrickom a Košickom kraji, na začiatku bola vysoká aj v Nitrianskom kraji. V Košickom kraji bola miera nezamestnanosti v rokoch 2000 až 2005 skoro „konštantná“, potom klesala, ale v roku 2010 vzrástla tak ako aj v celej Slovenskej republike. Bratislavský, Trenčiansky, Trnavský a Banskobystrický kraj mali počas sledovaného obdobia mieru nezamestnanosti pod celoslovenským priemerom. Miera nezamestnanosti v Nitrianskom kraji bola v rokoch 2000 až 2006 nad celoslovenským priemerom, ale potom sa jej hodnoty pohybovali okolo priemeru. Miera nezamestnanosti

v Žilinskom kraji kopíruje vývoj miery nezamestnanosti na Slovensku. Miera nezamestnanosti v Prešovskom, Košickom a Banskobystrickom kraji je počas celého sledovaného obdobia nad priemerom. Hodnoty aritmetického priemeru, smerodajnej odchýlky a variačného koeficientu pre dané roky sú vypočítané podľa vzťahov (1), (2), (3).



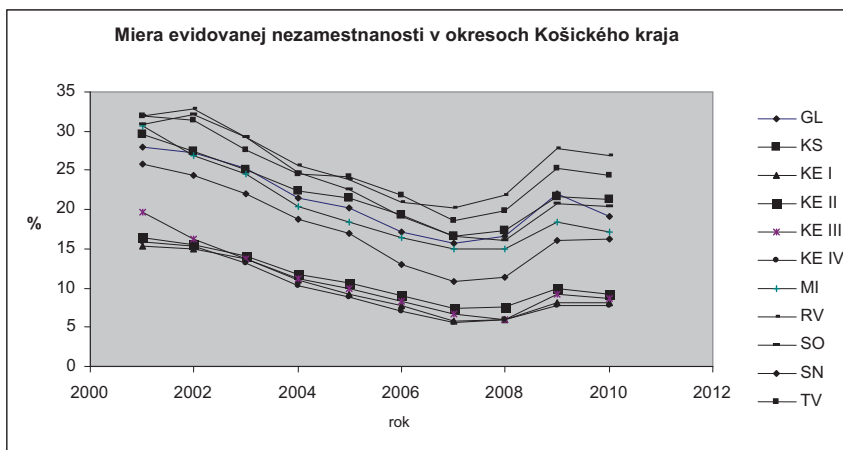
Graf č. 2: Vývoj miery nezamestnanosti v krajoch SR v r. 2000-2100

Z tabuľky vývoja evidovanej miery nezamestnanosti v rámci Košického kraja vyplýva, že počas sledovaného obdobia (2001-2010) najvyššia nezamestnanosť bola v okrese Rožňava a len v rokoch 2005 a 2006 bola vyššia v okrese Trebišov. Najnižšiu mieru nezamestnanosti mal na začiatku sledovaného obdobia okres Košice I, ale od roku 2003 až doteraz je to okres Košice IV, aj keď medzi jednotlivými okresmi sú veľmi malé rozdiely. Z grafu č. 3 vyplýva, že charakter rastu resp. poklesu miery evidovanej nezamestnanosti je vo všetkých okresoch podobný, líši sa iba hodnotami. V sledovanom období bola najvyššia miera evidovanej nezamestnanosti zaznamenaná v okrese Rožňava v roku 2002 (32,87 %). Najnižšiu mieru evidovanej nezamestnanosti dosiahol okres Košice IV v roku 2007 (5,54 %).

Tabuľka č. 3: Vývoj evidovanej miery nezamestnanosti v okresoch Košického kraja (v %) v r. 2001-2010.

okres/rok	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Gelnica	27,96	27,29	25,23	21,49	20,25	17,08	15,63	16,51	21,94	19,14
Košice - okolie	29,66	27,4	25,02	22,34	21,51	19,39	16,52	17,37	21,71	21,27
Košice I	15,39	14,99	13,68	11,01	9,23	7,8	5,85	6,01	8,09	8,07
Košice II	16,49	15,54	14,03	11,7	10,65	9,07	7,43	7,56	9,87	9,27
Košice III	19,61	16,19	13,76	11,11	9,9	8,38	6,66	6	9,22	8,59
Košice IV	15,82	15,27	13,15	10,33	8,83	7,05	5,54	5,89	7,82	7,82
Michalovce	30,59	26,85	24,5	20,46	18,46	16,48	15,01	14,95	18,32	17,21
Rožňava	31,96	32,87	29,27	25,58	23,77	21,01	20,14	21,74	27,75	26,82
Sobrance	30,9	32,11	29,19	24,76	22,49	19,19	16,59	16,09	20,66	20,34
Spišská N. Ves	25,79	24,35	21,98	18,72	16,94	12,95	10,77	11,43	16,14	16,28
Trebišov	31,89	31,47	27,64	24,6	24,23	21,84	18,6	19,8	25,24	24,42

Zdroj: portal.statistics.sk



Graf č. 3: Vývoj miery evidovanej nezamestnanosti v okr. Košického kraja v r. 2001-2010.

5. Záver

V predkladanom článku analyzujeme počet nezamestnaných a mieru evidovanej nezamestnanosti v Košickom kraji v porovnaní s ostatnými krajmi SR a taktiež v rámci kraja. Na základe údajov z portal.statistics.sk o nezamestnanosti sa nám podarilo zistiť, že v Košickom kraji je počet nezamestnaných vždy nad celoslovenským priemerom. Na porovnanie počtu nezamestnaných v Košickom kraji a v SR sme použili číselné charakteristiky a grafické nástroje štatistiky.

6. Literatúra

- [1] HINDLS, R., KAŇOKOVÁ, J.: Metódy statistickej analýzy pro ekonómy. Management Press, Praha 1997
- [2] Amird d ACZEL: Complete Business Statistics. Irwin, Boston 1989
- [3] <http://portal.statistics.sk/>
- [4] <http://px-web.statistics.sk/PXWebSlovak/index.htm>
- [5] <http://sk.wikipedia.org/wiki/Nezamestnanos%C5%A5>

Adresa autora:

Mgr. Gabriela Ižariková, PhD.
Letná 9
040 01 Košice
gabriela.izarikova@tuke.sk

RNDr. Denisa Olekšaková, PhD.
Letná 9
040 01 Košice
denisa.oleksakova@tuke.sk

Simulácia systému hromadnej obsluhy v programe R The Simulation of a Bulk Service System in Program R

Július Jenis, Jakub Jenis

Abstract: The article deals with programming language R and its usage in simulating bulk service theory problems.

Key words: Simulation, bulk service, language R

Kľúčové slová: Simulácia, hromadná obsluha, jazyk R

JEL classification: C15

1. Úvod

Predstavme si, že sa v reálnom živote stretne s nasledovným problémom.

Na dopravnom inšpektoriáte je k dispozícii 5 nezávislých obslužných miest (priehradok), doba obsluhy jedného vodiča pri priehradke sa riadi exponenciálnym rozdelením so strednou hodnotou 6 minút. Vodiči prichádzajú na úrad nezávisle, doba medzi príchodmi sa tiež riadi exponenciálnym rozdelením, ale priemerný počet prichádzajúcich vodičov (a teda aj stredná doba medzi príchodmi) sa v priebehu otváracích hodín úradu mení nasledujúcim spôsobom:

9:00 - 10:30 43 vodičov za hodinu;

10:30 - 15:30 28 vodičov za hodinu;

15:30 - 17:00 52 vodičov za hodinu.

V rovnakých časových intervaloch sú otvorené vždy 4, 3 resp. 5 priehradok. Ďalej predpokladajme, že žiadny vodič neodíde z úradu kvôli čakaniu vo fronte a že čakareň je dostatočne veľká (neexistuje teda žiadne obmedzenie pre príchody vodičov). Nakoniec ešte predpokladajme, že úrad obsluží všetkých vodičov aj po záverečnej, t.j. všetci vodiči, ktorí prišli pred 17:00 a o 17:00 stále čakajú, budú obslužení (aj po 17 hodine zostáva otvorených všetkých 5 priehradok).

Je to problém, ktorý s malými úpravami dosť reálne popisuje možnú situáciu bežného života. Je teda opodstatnené sa takýmto problémom zaoberať, obzvlášť vtedy, keď takýto problém je schopný generovať úlohy, ktorých riešenie spadá do oblasti teórie hromadnej obsluhy. Ide o vialinkový systém s čakaním v rade (bez strát). Majme teda za úlohu riešiť nasledovné úlohy. Zaujímajú nás príchody a obsluha vodičov na úrade počas 250 dní (jednotlivé dni sú medzi sebou nezávislé). Z tohto obdobia :

- Graficky znázorníte priemernú dobu čakania vo fronte v závislosti na čase príchodu vodiča (časy príchodov grupujte do 5 min. intervalov a uvažujte priemer pre všetky príchody v týchto intervaloch).
- Odhadnite pravdepodobnosť, že prichádzajúci vodič nebude musieť čakať vo fronte (neuvažujte závislosť na čase príchodu). Máte možnosť si vybrať dáta z ľubovoľného dňa alebo zo všetkých 250-tich dní. Čo si podľa svojho štatistického uváženia vyberiete (a prečo) ?
- Zostrojte histogram pre čas ukončenia obsluhy posledného vodiča. Porovnajte toto rozdelenie s normálnym.

2. Riešenie pomocou simulácie

Na simuláciu uvedeného problému sme použili program R. Je to voľne šíriteľný program, čo jeho použiteľnosť veľmi zvýrazňuje. Po krátkom štúdiu materiálov, ktorých je k dispozícii veľké množstvo (v anglickom, českom a aj slovenskom jazyku) sme sa mohli pustiť do prvých pokusov s generovaním potrebných údajov. Na tomto mieste by bolo vhodné spomenúť vhodnosť využívania tohto nástroja aj pri vyučovaní matematickej štatistiky a pravdepodobnosti na stredných školách. Takýmto spôsobom sa dá veľmi oživiť vyučovanie aj nie príliš jednoduchých, ale menej zaujímavých častí matematiky. Po zvládnutí a osvojení si potrebných funkcií daného jazyka potrebných na realizáciu nášho problému, sme už iba napísali program, ktorý to celé vygeneruje. Tieto údaje potom poslúžili ako vstupné údaje pre funkcie, ktoré dávali odpovede na naše otázky.

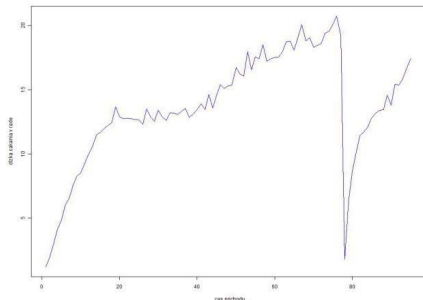
Úlohu riešime ako diskretnú simuláciu v programe R. Ako základnú časovú jednotku sme si zvolil jednu minútu. Simulujeme príchody a obsluhu vodičov na úrade podľa príslušných rozdelení. Základom simulácie je funkcia `simulacePrichody()`, ktorá vracia vektor príchodov jednotlivých vodičov za jeden deň. Generovanie prebieha v troch fázach, podľa toho, akým rozdelením sa práve príchody riadia. Nasleduje simulácia samotnej obsluhy počas dňa. Počet prichradok schopných obsluhy sa počas dňa priebežne mení. Po skončení simulácie máme väčšinu potrebných dát vo vektoroch:

`prichodyVektor` - `prichodyVektor[i]` je čas príchodu i -teho vodiča
`dobyObsluhy` - `dobyObsluhy[i]` je dĺžka obsluhy i -teho vodiča
`dobyCakania` - `dobyCakania[i]` je čas, ktorý i -ty vodič strávi vo fronte

Komentár k samotnej implementácii je v zdrojovom kóde, ďalej v texte budeme iba vysvetľovať postupy potrebné k samotnému získaniu štatistických dát. Nasleduje riešenie samotných úloh zo zadania.

A. Graficky znázorníte priemernú dobu čakania v závislosti na čase príchodu vodiča

Časy grupujeme do 5 minútových intervalov, tých bude 96, a vykreslíme graf závislosti doby čakania od príslušnosti k danému intervalu. Pre každý interval odhadujeme strednú hodnotu ako aritmetický priemer všetkých hodnôt, ktoré doňho spadajú. Priemerná doba čakania v závislosti na čase príchodu vodiča je zobrazená na obrázku 1.



Obrázok 1: Grafické znázornenie priemernej doby čakania.

Na grafe môžeme odpozorovať nasledovné veci. Na začiatku dňa sú fronty malé, priečinky však nestíhajú obsluhovať, takže postupne rastú. Okolo obeda sa znižuje počet prichádzajúcich vodičov, takže dĺžka čakania sa stabilizuje okolo 15 minút. Táto doba dramaticky klesá po otvorení ďalších dvoch priehradok o 15:30. Kvôli vyššiemu priemeru prichádzajúcich vodičov však postupne zase stúpa až po ukončenie dňa. Najviac sa oplatí prísť na začiatku dňa, alebo tesne po otvorení 5 priehradok.

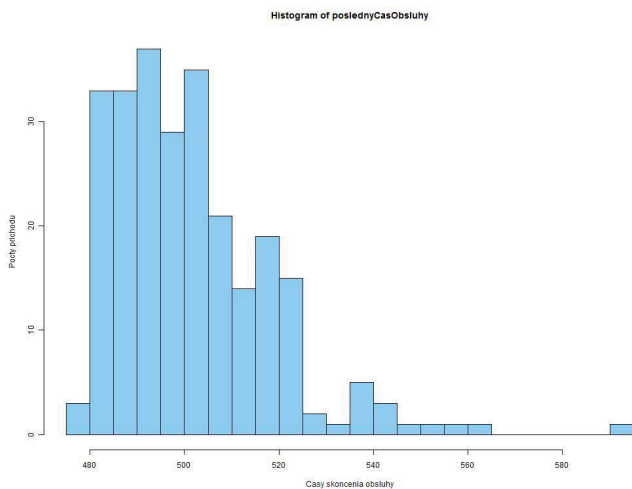
B Odhadnite pravdepodobnosť, že prichádzajúci vodič nebude musieť čakať

Presnejší odhad dostaneme, keď budeme pracovať s priemerom všetkých dní, aby sme minimalizovali výber extrémů. Výsledok je tak skôr priemerom, a viac odpovedá skutočnej pravdepodobnosti. K výsledku pridáme tak, že delíme počet vodičov, ktorý nečakali v rade celkovým počtom vodičov.

```
> pocetVodičovCelkem= 72082
> pocetVodicovNecakajucich= 12832
> pst=0.1780195 pravdepodobnosť, že vybraný vodič nečaká v rade
```

C Zostrojte histogram času obsluhy posledného vodiča. Porovnajete rozdelenie s normálnym rozdelením.

Pri zostrojení histogramu (obrázok 2) sme využili dáta z vektoru poslednyCasObsluhy.



Obrázok 2: Histogram času obsluhy posledného vodiča.

Ako test normality použijeme Shapiro-Wilkov test. Testujeme hypotézu: “hodnoty vektoru poslednyCasObsluhy sa riadia normálnym rozdelením” na hladine významnosti 0,05. Ak $P\text{-value} < 0,05$, zamietame hypotézu. Dostali sme výsledky testu:

```
> W = 0.9004,
> p-value = 5.989e-12
```

Na základe Shapiro-Wilkovho testu zamietame hypotézu. Čas ukončenia obsluhy posledného vodiča sa neriadi normálnym rozdelením.

Záver

V uvedenom príspevku sme sa snažili ukázať, ako je možné riešiť úlohy z reálneho života pomocou simulácie v programe R a na jednom sme demonštrovali konkrétne riešenie zadaných úloh.

3. Literatúra

[1]KRIŠŤÁK, M. 1998. Základy teórie stochastických procesov. AX INZERT, Bratislava.

[2]KOMÁREK, A. 2010. Hrátky s R, 197, MFF KU Praha4.

[3][HTTP://WWW.R-PROJECT.ORG/](http://www.r-project.org/) (INFORMACE, MANUÁLY, FAQ)

Adresy autorov:

Július Jenis, RNDr.

Tr.A.Hlinku č.1

949 01 Nitra

jjenis@ukf.sk

Jakub Jenis

Pod orešinou 19

949 11 Nitra

jack_jenis@yahoo.com

Štatistická gramotnosť v súvislosti s polopravdami a zavádzajúcimi vyjadreniami v štatistike

Statistical literacy in connection with half-truths and confusing statements in statistics

Mária Kóšová

Abstract: In today's society full of statistical statements, everybody should be able to orient themselves in these statements. It means that one should be statistically literate. This article presents what the term statistical literacy implies. It deals with the ability to think critically about statistical statements and through it detect various inaccuracies, confusing statements and half-truths, whether intentional or unintentional.

Key words: Statistical literacy, Confusing statements in statistics, Half-truths in statistics

Kľúčové slová: Štatistická gramotnosť, máťúce tvrdenia v štatistike, poopravdy v štatistike

JEL classification: C19

1. Úvod

V súčasnej dynamicky sa rozvíjajúcej informačnej spoločnosti čoraz viac narastá význam štatistiky. Každý deň sme totiž zaplavení rôznymi štatistickými informáciami. Napr. reklama („9 z 10 lekárov odporúča ...“), spravodajstvo („prieskum verejnej mienky ukazuje, že ...“), rôzne tabuľky a grafy v rukách politikov, ktorí sa nás nimi snažia presvedčiť o raste resp. poklese cien, inflácie a i. za ich vlády, prípadne nimi podporujú svoje tvrdenia o potrebnosti schválenia niektorého zákona, rôzne trendy a javy v spoločnosti (kriminalita, populačný rast, zamestnanosť, rozšírenosť chorôb, úspešnosť tej ktorej liečby, ...). Rovnako tak odborníci a hlavne advokáti používajú štatistické a numerické vyjadrenia, aby posilnili svoje argumenty. Mnohokrát sú však tieto informácie použité k prekrúcaniu alebo zahmlievaniu skutočných údajov. Práve v spojení s týmto sa čoraz viac dostáva do popredia štatistická gramotnosť. Pretože len štatisticky gramotný človek sa v týchto tvrdeniach a v rôznych štatistických informáciách dokáže orientovať a kriticky nad nimi uvažovať. V neposlednom rade štatistická gramotnosť pomáha sa správne rozhodovať aj v situáciách, ktoré sú založené na princípe náhody (napr. lotto, poisťné udalosti a zmluvy, ...).

2. Čo je to štatistická gramotnosť

Štatistickú gramotnosť možno definovať ako schopnosť pracovať so štatistickými informáciami napr. vo forme rôznych vyjadrení, výskumov, tabuľkách alebo grafoch. V rámci toho možno povedať, že štatisticky gramotný človek by mal byť schopný čítať štatistické informácie, interpretovať ich, to znamená vidieť v štatistických údajoch čísla v kontexte v ktorom sa nachádzajú, a hlavne kriticky uvažovať o týchto číslach a interpretáciách, ktoré sú mnohokrát použité ako dôkaz podporujúci nejaké tvrdenie.

Milo Schield rozoznáva tri oblasti štatistickej gramotnosti:

1. založená na náhode (chance - based)
2. založená na klame (fallacy – based)
3. založená na korelácii (correlation – based)

Oblasť štatistickej gramotnosti založená na náhode

Zaoberá sa predovšetkým variabilitou spôsobenou náhodou – náhodná variabilita. To znamená napr. štatisticky významné vzťahy sa môžu stať štatisticky nevýznamnými, ak vezmeme do úvahy vplyv ďalšieho súvisiaceho faktora stať štatisticky nevýznamnými (na predaj množstva zmrzlina a počet nehôd v meste vplýva počet obyvateľov), pravdepodobnosť zriedkavých udalostí (100 ročná voda, najmenšie zrážky, ...), vzťah medzi rozsahom výberu a intervalmi spoľahlivosti, rozdiel medzi 95% intervalom spoľahlivosti a 95% pravdepodobnosťou vytiahnutia bielej guľky z vrečka, testovanie hypotéz (možno povedať, že čím menšia je p - hodnota tým viac svedčí o platnosti alternatívnej hypotézy?) a i.

Oblasť štatistickej gramotnosti založená na korelácii

Obsahuje v sebe oblasti ako percentá – čo je celok a čo je časť (25% dievčat je z Česka vs. 25% Čechov sú dievčatá), čítanie a interpretovanie tabuliek a grafov (všetky samovraždy tvoria 100%, z nich je percento vdov spáchajúcich samovraždu väčšie ako percento vdovcov spáchajúcich samovraždu, avšak, pretože dané percentá nezohľadňujú počet vdov a vdovcov, táto rozdielnosť v počte percent nemusí znamenať, že ženy (vdovy) nesú stratu manžela ťažšie a teda sú náchylnejšie spáchať samovraždu ako muži (vdovci) ([5])), podmienená pravdepodobnosť, rozdiel medzi vzťahom a príčinnosťou a i.

Oblasť štatistickej gramotnosti založená na klame

Zahrňa v sebe štatistickú gramotnosť v zmysle kritického uvažovania o štatistických vyjadreniach. Teda rozoznávanie pravdivých tvrdení, poloprávd a zavádzajúcich tvrdení, nepravdivých tvrdení, ktoré sú predkladané a v rámci uvažovania nad týmito závermi kladenie správnych otázok nabádajúcich k zisteniu podmienok pravdivosti vyjadrenia.

V dnešnej dobe sa na nás hlavne z médií dennodenne valia rôzne grafy, tabuľky, výsledky ankiety, prieskumov, reklám, ktoré požívajú rôzne štatistiky ako dôkazy nejakých tvrdení alebo úžasnosti ich výrobkov. Práve preto možno kritické uvažovanie nad štatistickými vyjadreniami považovať za nevyhnutnú schopnosť, ktorú by mal mať každý človek, aby dokázal odhaliť, čo je klamstvo, pravda alebo tzv. polopravda prípadne zavádzanie a na základe toho sa správne rozhodovať v bežných záležitostiach života ako sú voľby, trendy v spoločnosti, dôveryhodnosť reklamy, odborné rady, lekárske odporúčania a pod.

Preto sa v ďalšom budeme zaoberať niektorými spôsobmi prezentovania tzv. poloprávd prípadne zavádzania prostredníctvom štatistických tvrdení - či už úmyselného alebo neúmyselného

3. Polopravdy a zavádzania v štatistike

Pre lepšiu predstavu zmyslu pomenovania „polopravda“ uvedieme nasledovný príbeh zo života:

V škole udrie Janko spolužiaka, ktorý mu úder vráti. Pani učiteľka to videla a Jankovi napísala poznámku: „Váš syn sa bil so spolužiakom“. Otec sa ho doma opýta, že čo sa stalo a Janko mu odpovie: „, Spolužiak ma udrel“. Klamal Janko svojmu otcovi? Neklamal, avšak povedal len polopravdu. Janko totiž zamľčal, že bitku začal on.

Ako vidno z príbehu alebo ako intuitívne cítime kľúčkovanie pomocou štatistiky (polopravdy v štatistike) je rozdielne od klamaní pomocou štatistiky. Rovnako tak inou kategóriou sú tiež čitateľské chyby. Čitateľskými chybami rozumieme vlastné chyby čitateľa spôsobené nedostatočnými matematickými vedomosťami alebo nepozorným čítaním úplnej, presnej a jednoznačnej štatistickej informácie. Polopravdy vznikajú úmyselným alebo

neúmyselným zmlčaním niektorého dôležitého údaju alebo prekrútením skutočných výsledkov. Čím pri interpretácii konečným prijímateľom vzniká veľký priestor na nejednotný výklad. Vďaka tomu, často aj my samotní označujeme mnohé štatistické vyjadrenia a výsledky prieskumov za klamstvá alebo ich naopak nekriticky prijímame avšak veľa krát ide „len“ o polopravdy. Tento unáhlený názor môže byť spôsobený tým, že si pri spracovávaní významu vyjadrenia mnohé veci či už svojvoľne domyslíme alebo sa jednoducho nad niektorými skutočnosťami nezamýšľame.

Jedným z mnohých zdrojov nekorektných vyjadrení sú rôzne prieskumy a ankety. Dôvodom môže byť spôsob kladenia otázok. Nedávno sa stal takýto prípad v Bratislave, kde boli aj kvôli tendenčne kladeným otázkam (t.z. navádzajúcim na želanú odpoveď) spochybnené výsledky prieskumu spoločnosti Metro Bratislava týkajúce sa záujmu obyvateľov o električku v mestskej časti (MsČ) Petržalka. Uvedieme jednu z otázok prieskumu:

„Cesta električkou je približne o tretinu rýchlejšia ako autobusom. Keď budete mať možnosť cestovať z Petržalky električkou, využili by ste túto možnosť MHD?“ Áno Nie

Je jasne vidieť, že otázka je formulovaná tak, že navádza na kladnú odpoveď. Iným spôsobom kladení tendenčných otázok môže byť škála ponúkaných odpovedí napr.

Pracovník mi bol vždy ochotný poradiť.

Silno súhlasím Úplne súhlasím Súhlasím Neviem Nesúhlasím

Vzhľadom nato, že škála ponúka tri kladné odpovede a len jednu zápornú je vidieť, že nabáda odpovedať kladne. Samozrejme výsledky takéhoto prieskumu alebo ankety nie sú relevantné. S prieskumom alebo anketou súvisí aj mätúca interpretácia ich výsledkov. Nasledovný príklad ilustruje jednu z možností takejto interpretácie:

V nitrianskom kraji bol uskutočnený prieskum, ktorého cieľom bolo zistiť, či sú vodiči spokojní s prácou cestárov v zimnom období. Výsledky hovoria, že až 65% percent opýtaných si myslí, že cestári túto zimu zvládli na 8 bodov z celkovej 10 bodovej škály, pričom 10 bodov predstavovalo najväčšiu spokojnosť. Avšak v celoslovensky distribuovaných novinách sa objaví titulok: „Podľa prieskumu zvládli cestári minulú zimu na 8 bodov z 10-tich“. Je táto informácia klamstvo? Je to pravda? Odpoveďou je, že je to polopravda. Titulok totiž neprávom zovšeobecňuje výsledky Nitrianskeho kraja pre celé Slovensko. Mnohí čitatelia nekriticky prijímú túto informáciu bez akýchkoľvek pochybností a mnohí ju zavrhnú ako klamstvo. Avšak len štatisticky gramotný čitateľ najskôr hľadá odpoveď na otázku: „Kde bol uskutočnený tento prieskum, aký bol výber respondentov a koľko percent respondentov si to myslí?“ a až na základe odpovedí na ne si interpretuje daný titulok. Výsledky a závery niektorých prieskumov sú často buď zámerne alebo nechtiac zle prezentované prípadne nami pochopené ako výsledky platné pre celú populáciu. Podobne je to aj s rôznymi politickými reláciami, ktoré umožňujú zavolať alebo poslať SMS ako vyjadrenie sympatie resp. antisympatie voči prítomnému politikovi. Konečný výsledok hlasovania vytvára u mnohých divákov mylnú predstavu o obľúbenosti resp. neobľúbenosti politika v rámci celého obyvateľstva. Táto chyba v prezentácii alebo interpretácii výsledkov prieskumu či ankety spočíva v nereprezentatívnom výbere.

Ďalším a určite nie zanedbateľným zdrojom poloprávď resp. nepresných a zavádzajúcich vyjadrení sú reklamy, ktoré využívajú štatistické vyjadrenia ako podporu predajnosti svojho výrobku. Napr. presnosť tehotenského testu. Väčšina výrobcov sa snaží predajnosť podporovať vyjadreniami o čo najlepšej presnosti testu. Niektorí uvádzajú presnosť až 99,9% ale väčšinou sa v rôznych reklamách hovorí o presnosti 97% - 99%. Vezmime si napr. presnosť 99%. Čo táto informácia vlastne vyjadruje? 99% žien s pozitívnym výsledkom sú skutočne tehotné alebo u 99% tehotných žien ukáže test pozitívny výsledok? Nieкто by si mohol myslieť, že presnosť 99% znamená, že 99% žien, ktorým ukáže

test pozitívny výsledok sú skutočne tehotné a 1% žien, ktorým ukáže test pozitívny výsledok nie sú tehotné (falošná pozitívnosť). Avšak, ak sa nad touto interpretáciou zamyslíme, zistíme, že to by ale znamenalo, že ak otestujeme skupinku žien z ktorých je veľmi nízky počet tehotných (napr. 1 zo 100), potom 50% žien s pozitívnym výsledkom testu by bolo falošne pozitívnych. Ako je to možné pri tak veľkej presnosti? Pretože presnosť 99% znamená presnosť v potvrdení nejakej skutočnosti t.z. že 99% tých žien, ktoré sú skutočne tehotné ukáže test aj pozitívny výsledok. Prvé vysvetlenie presnosti je „presnosť v predpovedaní a tá ako sme ukázali určite nie je 99%. Preto samotné vyjadrenie „presnosť X%“ je zavádzajúce a môžeme ho teda označiť ako polopravda.

Nemenej zaujímavým a bohatým zdrojom zavádzania je aj podmienená pravdepodobnosť, teda Bayesov vzorec. Tento používame mnohokrát aj my v bežnom živote. Každému z nás sa určite stalo, že musel kvôli zaneprázdnenosti na poslednú chvíľu odmietnuť stretnutie s priateľom či priateľkou. Avšak potom zrazu tento priateľ príliš často nedvíha telefón a ak dvihne, tvrdí, že má veľa práce. Napriek tomu sme ho videli na káve v meste. Ale my nevieme, že v práci naozaj robí nejaký nový projekt a aj spomínaná káva bola pracovná. Takmer všetci by sme hneď usúdili, že je určite urazený kvôli nášmu odmietnutiu. Teda, že je oveľa väčšia pravdepodobnosť, že sa takto správa, ak je urazený ako keď nie je. Ale práve takýmto uvažovaním sa dopúšťame chyby. Zamenili sme totiž pravdepodobnosť, že je priateľ urazený, ak sa takto správa s pravdepodobnosťou, že sa takto správa, ak je urazený. Práve z tohto omylu, mylnej predstavy o možnosti zamenenia daných pravdepodobností, prameni podstata zavádzania pomocou podmienenej pravdepodobnosti. Uvedený nedostatok v používaní podmienenej pravdepodobnosti môže mať oveľa závažnejší rozmer ak sa ho dopustíme my alebo nejaký lekár pri interpretovaní prognóz závažného ochorenia na základe výsledkov testu. Test HIV dáva pozitívny výsledok aj keď krv nie je nakazená v 1 z 1000 krvných testov. Chyba v interpretácii môže nastať, ak výsledok pozitívneho testu formulujeme ako pravdepodobnosť 999 ku 1000, že je dotýčaný nakazený HIV a v priebehu pár rokov zomrie. Zamenili sme totiž pravdepodobnosť, že nie je nakazený, ak má pozitívny test s pravdepodobnosťou, že má pozitívny test, ak nie je nakazený.

Takéto zámery mali v minulosti negatívny dopad v americkom súdnictve. Vzhľadom nato, že v súdnych sporoch je porota zložená z bežných občanov, nachádzajú tu veľké uplatnenie zavádzajúce, na prvý pohľad pravdivé a logické, argumenty a tvrdenia (označované aj ako žalobcov klam). Takáto nesprávna argumentácia v súvislosti s porotou, ktorá ju nebola schopná odhaliť, mala dokonca za následok justičné omyly. Jedným z nich bol prípad Sally Clarkovej. Jej prvé dieťa zomrelo vo veku 11 týždňov. Jeho smrť bola odôvodnená syndrómom náhleho úmrtia dojčiat – SIDS. Sally však mala potom aj druhé dieťa, ktoré tento krát zomrelo po ôsmich týždňoch, údajne tiež na SIDS. Ale na základe úmrtia druhého dieťaťa bola Sally obvinená z dvojnásobnej vraždy zadusením. V prospech obžaloby vtedy svedčil Roy Meadow, expert na detské lekárstvo. Vypovedal, že vzhľadom na veľmi vzácny výskyt SIDS je pravdepodobnosť úmrtia oboch detí v dôsledku tohto syndrómu 1 ku 73 miliónom. A práve toto svedectvo presvedčilo porotu, že Sally je vinná. Roy Meadow odhadol, že pravdepodobnosť úmrtia dieťaťa na SIDS je 1 ku 8543. Pravdepodobnosť 1 ku 73 miliónom získal vynásobením pravdepodobnosti úmrtia každého dieťaťa zvlášť. Avšak týmto vychádzal z predpokladu, že tieto dva javy sú nezávislé t.z. že neexistuje žiadna genetická dispozícia alebo iný vplyv, ktorý by zvyšoval pravdepodobnosť úmrtia druhého dieťaťa potom, čo v dôsledku tohto syndrómu zomrelo prvé dieťa. Práve v takejto úvahe spočívala chyba vo výpočte pravdepodobnosti. V tomto prípade nešlo o pravdepodobnosť, že dve deti (nezávisle na sebe) zomrú v dôsledku SIDS, ako to prezentoval žalobca a Roy Meadow, ale o pravdepodobnosť úmrtia dvoch detí v dôsledku SIDS, ak naozaj zomreli. Podobným prípadom bol aj prípad O.J. Simpsona. Avšak teraz pomohol tzv. žalobcov klam k jeho oslobodeniu aj napriek tomu, že existovali relevantné dôkazy o jeho vinne (pozri [3]).

4. Záver

V našom príspevku sme uviedli len niektoré možnosti a spôsoby zavádzania a predkladania tzv. poloprávď v štatistike. Ale ako možno vidieť z predchádzajúceho, byť štatisticky gramotný, to znamená okrem iného aj kriticky uvažovať a porozumieť štatistickým vyjadreniam, sa stáva v dnešnej spoločnosti nevyhnutnosťou. Práve preto sa prostredníctvom školskej reformy dostáva štatistická gramotnosť viac aj do škôl. Bolo urobených viacero zmien týkajúcich sa práve širšieho zaoberania sa štatistikou už v základných školách. Pri vyučovaní je kladený dôraz na úlohy s kontextom s reálneho života, čím sa dá budovať štatisticky gramotnejšia spoločnosť schopná robiť aj na základe štatistiky správne rozhodnutia.

5. Literatúra

- [1]BEST, J. 2004. More Damned Lies and Statistics – How Numbers Confuse Public Issues. California: University of California press, 2004. 1076 s. ISBN 0-520-3-3-3.
- [2]BOLTON, P. Statistical literacy. Dostupné na internete (28.4.2011): <http://www.parliament.uk/briefingpapers/commons/lib/research/briefings/snsg-05708.pdf>
- [3]MLODINOW, L. 2008. Život je jen náhoda – jak náhoda ovlivňuje naše životy. Praha: Slovart, s. r. o., 2009. 246 s. ISBN 978-80-7391-259-8
- [4]SCHIELD, M. 1999. Statistical Literacy: Thinking critically about statistics. Dostupné na internete (27.4.2011): <http://www.statlit.org/pdf/1999SchieldAPDU.pdf>
- [5]SCHIELD, M. 2001. Three kinds of statistical literacy: what should we teach?. Dostupné na internete (28.4.2011): <http://www.statlit.org/PDF/2002SchieldICOTS.pdf>
- [6]SCHIELD, M. 2005. Statistical prevarication: Telling half truths using statistics. Dostupné na internete (26.4.2011): <http://www.statlit.org/pdf/2005SchieldIASE.pdf>
- [7]SME: Apríl 2011. Dostupné na internete (21.4.2011): <http://bratislava.sme.sk/c/5850043/uhler-vysledky-prieskumu-o-elektricke-v-petrzalke-nie-su-smerodajne.html>

Adresa autora:

Mária Kóšová, Mgr.
Univerzita Konštantína Filozofa v Nitre
Tr. A. Hlinku 1
949 74 Nitra
maria.kosova@ukf.sk

Použitie Kruskalovho-Wallisovho testu pri overovaní účinnosti kombinovaného spôsobu vyučovania anglického jazyka
The Application of the Kruskal-Wallis' Test in the Testing of Efficiency of Combined Methods of Teaching of Professional English

Zuzana Kurucová, Beáta Stehlíková, Anna Tirpáková

Abstract: The presented article deals with the basic aspects of teaching English as a foreign language in three different forms: the classic (in-class) model, the e-learning model and the combined model (combining classic teaching/learning and e-learning). The test used to confirm or reject the basic presumptions of the experiment was the Kruskal-Wallis' test. The test was used in order to evaluate the quality of the combined method and its efficiency.

Key words: investigation, Kruskal-Wallis' test, methods of teaching, combined method of teaching, e-learning, classic teaching, evaluating the efficiency of a new method of teaching

Kľúčové slová: experiment, Kruskalov-Wallisov test, vyučovacie metódy, kombinovaná metóda vyučovania, e-learning, klasické vyučovanie, overovanie účinnosti nového spôsobu vyučovania

JEL classification: C18

1. Úvod

V súčasnosti, keď sa rozšírili možnosti uplatniť sa na trhu práce v rámci celej Európy, tak znalosť cudzích jazykov sa stala nevyhnutnosťou pre každého človeka, ktorý hľadá uplatnenie vo svojom odbore. Príprava na budúce povolanie už nezahŕňa len klasickú výučbu v triedach, ale čoraz väčší priestor je venovaný alternatívnym modelom výučby. Medzi takéto alternatívne modely, ktoré dopĺňajú a v niektorých prípadoch aj nahrádzajú klasické výučbové prístupy, patrí aj e-learning. E-learning je vo svojej podstate virtuálnou formou vzdelávania, ktorá môže byť za určitých okolností buď úplne alebo čiastočne dopĺňovaná klasickými formami výučby prostredníctvom pedagóga a v priamej (nie reprodukovanej komunikácii) s ním. Ak skúmame tzv. klasické e-learningové modely vzdelávania, kde počítač chápeme v užšom zmysle (t.j. ako hardware), potom môžeme v tomto systéme identifikovať najmä nasledovné základy submoduly e-learningu:

- a) profesionálny software (interaktívne CD, DVD a pod., ktoré sa neviažu na prístup k internetu),
- b) on-line vzdelávanie prostredníctvom internetu a na to určených programov, resp. web stránok,
- c) on-line vzdelávanie prostredníctvom internetu, avšak bez web stránok na to priamo určených (napr. e-mailové konzultácie, e-mailové zasielanie dokumentov, úloh a pod.).

Vzhľadom na to, že v podmienkach Slovenskej republiky e-learning v súčasnosti nemá podstatné postavenie a má viac doplnkový ako nahrádzajúci charakter, bolo našou snahou v praxi overiť teóriu predpokladané výhody e-learningu. Experiment bol realizovaný na troch skupinách študentov: študenti, ktorí sa vzdelávali len v klasickom modeli výučby (Standard), ďalej študenti, ktorí sa vzdelávali len v režim on-line (On line) a tretiu skupinu tvorili študenti, ktorí navštevovali „klasické“ hodiny a v doplnkovom postavení aplikovali a vzdelávali sa prostriedkami e-learning (Combined). Výučba bola zameraná na odbornú angličtinu – English in the Media, t.j. na odborný jazyk používaný v masmédiách a pre potreby osôb pôsobiacich v tejto oblasti.

V predložennom príspevku analyzujeme výsledky, ktoré dosiahli respondenti, zaradení do v jednotlivých skupín. Pre vyhodnotenie výsledkov výskumu pomocou štatistických metód

sme použili okrem opisnej štatistiky aj neparametrické testovacie metódy, a to Kruskalov-Wallisov test. Výpočty boli realizované pomocou programu STATISTICA

2. Materiál a metódy

Študenti (členovia) jednotlivých skupín boli vyberaní náhodným výberom – na základe vlastného uváženia a prihlásenia sa do projektu. Na základe uvedeného boli vytvorené tri heterogénne skupiny, ktorých rôznorodosť mala znalostný charakter (v skupine boli študenti s jazykovými znalosťami B2 až C1 podľa Spoločného európskeho rámca) vo vekuom rozpätí 20 až 55 rokov, ženy aj muži.

Na začiatku projektu prebehol tzv. pretesting, t.j. overenie jazykových znalostí členov všetkých troch skupín pred začatím experimentu a tzv. posttesting, t.j. overenie jazykových znalostí členov všetkých troch skupín po skončení experimentu. Pretesty a posttesty sa zameriavali na overenie tých istých zručností, aby tieto boli kvantifikovateľné a aby mohla byť uskutočnená komparácia.

Testy overujúce písomnú znalosť (Writing) sa skladali z cvičení nasledovného typu: Grammar (G), Terminology (T) a Structure (S). Overovali sa teda schopnosti študentov tvoriť koherentný text, ktorý je gramaticky správny a používa sa v ňom vhodná terminológia. Za každú z uvedených položiek mohol študent získať maximálne 100 bodov (100%), pričom celkovo za túto časť teda mohol získať 300 bodov.

Overovanie znalostí študentov v časti „Listening“ sa zameriavalo na nasledovné subsystémy jazyka a vnímania: Reproduction (R), Questioning (Q) a Terminology (T). Testom sa zisťovali, či študent pochopil text ako celok (reproduction), či bol schopný vnímať konkrétosti (odpovedať na konkrétne otázky pedagóga) a či v rámci odpovedí používa správnu terminológiu (T). Za každú z uvedených položiek mohol študent získať maximálne 100 bodov (100%), pričom celkovo za túto časť teda mohol získať 300 bodov.

Testová časť v sekcii „Speaking“ overovala plynulosť rečových – jazykových schopností (Fluency = F), vhodnosť použitej slovnej zásoby (Terminology = T) a primeranosť gramatických väzieb (Grammar = G). Za každú z uvedených položiek mohol študent získať maximálne 100 bodov (100%), pričom celkovo za túto časť teda mohol získať 300 bodov.

Slovná zásoba ako posledná zložka testu „Vocabulary“ skúmala znalosť všeobecnej slovnej zásoby (Universal = U), odbornej slovnej zásoby (Professional = P) a vhodnosť ich aplikácie (Application = A). Za každú z uvedených položiek mohol študent získať maximálne 100 bodov (100%), pričom celkovo za túto časť teda mohol získať 300 bodov.

Z uvedeného vyplýva, že študent mohol v rámci pretestu resp. posttestu získať celkovo najviac 1.200 bodov (4 sekcie po 300 bodov). Pri vyhodnocovaní záverov - testíngov sme aplikovali tzv. Kruskalov - Wallisov test, ktorého podstatu opisujeme nižšie.

Kruskalov - Wallisov test je neparametrickou obdobou jednofaktorovej analýzy rozptylu, t.j. umožňuje testovať hypotézu H_0 , že k ($k \geq 3$) nezávislých súborov pochádza z toho istého rozdelenia. Je priamym zovšeobecnením Wilcoxonovho dvojitýberového testu pre prípad k nezávislých výberových súborov ($k \geq 3$).

$$K = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{T_i^2}{n_i} - 3(n+1),$$

ktorá má asymptoticky χ^2 - rozdelenie s $k-1$ stupňami voľnosti.

Testovanú hypotézu H_0 zamietame na hladine významnosti α , ak $K \geq \chi_\alpha^2(k-1)$, kde $\chi_\alpha^2(k-1)$ je kritická hodnota χ^2 - rozdelenia s $k-1$ stupňami voľnosti. Ak zamietneme testovanú hypotézu H_0 v prospech alternatívnej hypotézy H_1 , ktorá znamená, že ošetrenia neprinášajú rovnaký efekt, ostáva nezodpovedaná otázka, ktoré výbery sa od seba štatisticky

významne líšia. V analýze rozptylu sa na zodpovedanie tejto otázky používa Duncanov test, Tukeyho metóda, Scheffeho metóda alebo Neményiho test. V našom prípade sme mnohonásobné porovnanie poradových priemerov pri Kruskal – Wallisovom teste v programe STATISTICA realizovali Kruskalovým - Wallisovým testom viacnásobného porovnania pomocou Z – skóre

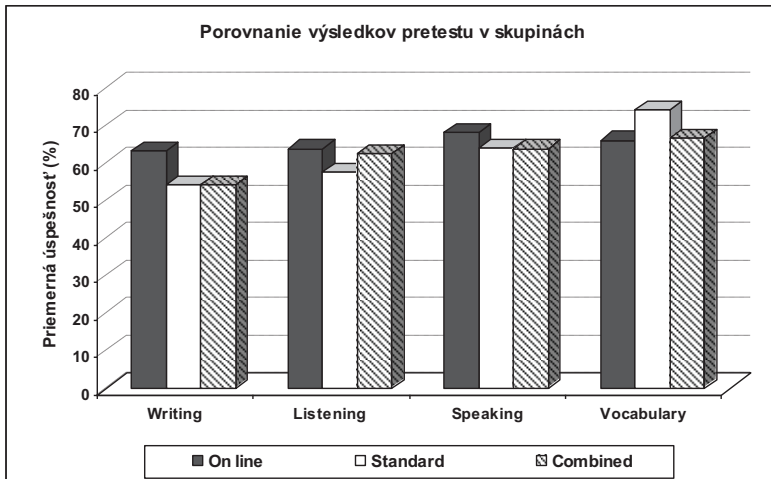
$$Z = \frac{|\bar{R}_i - \bar{R}_j|}{\sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}; \quad Z \geq z_\alpha$$

Vo výstupnej zostave jedna tabuľka udáva Z – skóre a druhá príslušné p – hodnoty.

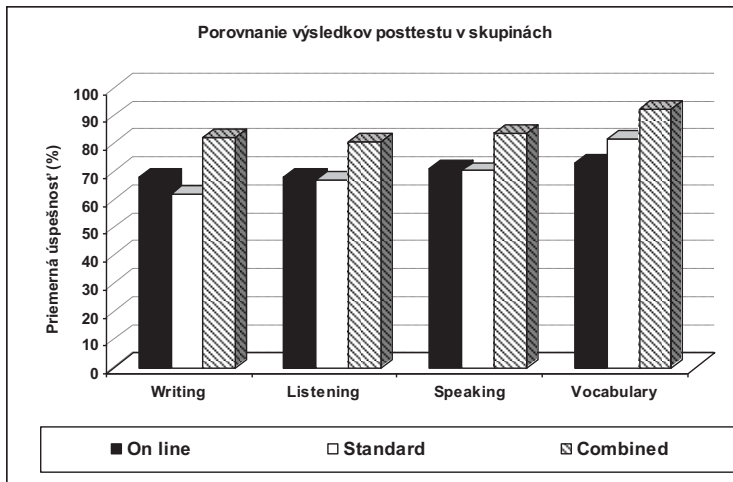
3. Výsledky a diskusia

Výsledky experimentu boli spracované najskôr pomocou metód popisnej štatistiky. Pre ilustráciu uvádzame grafické vyhodnotenie výsledkov pretestu a posttestu v jednotlivých skupinách.

Obrázok 1: Výsledky pretestu v skupinách respondentov z jednotlivých oblastí



Obrázok 2: Výsledky posttestu pretestu v skupinách respondentov z jednotlivých oblastí



Na obrázkoch 1 a 2 vidíme, že medzi skupinami v dosiahnutých výsledkoch v predteste a postteste sú rozdiely. Pomocou štatistických metód chceme zistiť, ktoré s uvedených rozdielov sú štatisticky významné, t.j. či úroveň faktora, ktorý predstavuje jednotlivé skupiny respondentov, má vplyv na priemernú hodnotu pozorovaného znaku. Ak počet porovnávaných súborov (počet úrovní faktora) je k , potom problém môže byť sformulovaný v tvare štatistických hypotéz nasledovne. Testovanou hypotézou je hypotéza

H_0 : rozdelenia k základných súborov sú identické. Oproti testovanej hypotéze kladieme alternatívnu hypotézu H_1 : nie všetkých k rozdelení je identických. Ekvivalentne môže byť testovaný problém sformulovaný nasledovne H_0 : priemerné hodnoty pozorovaného znaku nezávisia od úrovni faktora, H_1 : priemerné hodnoty závisia od úrovni faktora.

Keďže nie je opodstatnený predpoklad o normálnom rozdelení pozorovaných znakov a porovnáваме niekoľko základných súborov na základe závislých výberových súborov, na testovanie nulovej hypotézy bol použitý neparametrický Kruskal-Wallisov test.

Kruskalovým – Wallisovým testom sme overovali, či sa tri podskupiny vytvorené podľa úrovni faktora - typ skupiny (On line, Standard, Combined) štatisticky významne líšia v pozorovanom znaku – úspešnosť v preteste (= počet bodov). Teda $k=3$, pričom $n_1 = 20$, $n_2 = 20$, $n_3 = 12$, $n = n_1 + n_2 + n_3 = 32$ respondentov.

Kruskal – Wallisov test sme realizovali pomocou programu STATISTICA. Po zadaní vstupných údajov vo výstupnej zostave počítača dostaneme pre zvolený Kruskal – Wallisov test tieto výsledky: hodnotu testovacieho kritéria H a hodnotu pravdepodobnosti p , čo je pravdepodobnosť chyby, ktorej sa dopustíme, keď zamietneme testovanú hypotézu. Ak je

vypočítaná hodnota p dostatočne malá ($p < 0,05$ resp. $p < 0,01$), testovanú hypotézu H_0 o rovnosti stredných úrovní pozorovaných znakov zamietame (na hladine významnosti 0,05 resp. 0,01). V opačnom prípade hypotézu H_0 nemôžeme zamietnuť, t.j. pozorované rozdiely nie sú štatisticky významné

Po zadaní výsledkov pretestu sme použitím Kruskalovho-Wallisovho testu vypočítali hodnotu testovacieho kritéria $H = 0,433566$ a hodnotu pravdepodobnosti $p = 0,8051$. Keďže vypočítaná hodnota pravdepodobnosti p je väčšia ako 0,05, nulovú hypotézu H_0 nemôžeme zamietnuť na hladine významnosti $\alpha = 0,05$. To znamená, že skupiny sa štatisticky významne nelíšia vo výsledkoch pretestu. Všetky tri skupiny dosiahli približne rovnakú úspešnosť vo výsledkoch pretestu.

Rovnakým spôsobom boli vyhodnotené výsledky posttestu. Použitím Kruskalovho-Wallisovho testu vypočítali hodnotu testovacieho kritéria $H = 11,27376$ a hodnotu pravdepodobnosti $p = 0,0036$. Keďže vypočítaná hodnota pravdepodobnosti p je menšia ako 0,05, nulovú hypotézu H_0 zamietame na hladine významnosti 0,01 v prospech alternatívnej hypotézy. Testom sa potvrdilo, že jednotlivé skupiny respondentov (On line, Standard, Combined) sa štatisticky významne líšia vzhľadom na úspešnosť v postteste.

Vzhľadom na to, že testom sa potvrdilo, že medzi pozorovanými skupinami je štatisticky významný rozdiel, zaujímalo nás, ktoré skupiny sú vzájomne štatisticky významne odlišné. Na túto otázku nám dal odpoveď Kruskal - Wallisov test viacnásobného porovnávania, ktorý sme realizovali v programe STATISTICA. Vo výstupnej zostave z počítača sme použitím Kruskal - Wallisovho testu viacnásobného porovnania dostali hodnoty Z - skóre a príslušné hodnoty pravdepodobností p . Pretože testovanú hypotézu H_0 zamietame (na hladine významnosti 0,05 resp. 0,01), ak je vypočítaná hodnota p dostatočne malá ($p < 0,05$ resp. $p < 0,01$) a H_0 nemôžeme zamietnuť, ak je vypočítaná hodnota $p \geq 0,05$, z výstupnej zostavy uvádzame iba tabuľky s vypočítanými hodnotami pravdepodobností p .

Tabuľka 1: Výsledky Kruskal - Wallisovho testu viacnásobného porovnania v postteste

	On line	Standard	Combined
On line		0,997000	0,015937*
Standard	0,997000		0,004079*
Combined	0,015937*	0,004079*	

Z tabuľky vidíme, že vzhľadom na výsledky v postteste sa štatisticky významne líšia medzi skupinami On line a Combined ($p = \mathbf{0,015937^*}$) a tiež Standard a Combined ($p = \mathbf{0,004079^*}$), ale medzi skupinami On line a Standard ($p = 0,997000$) nie je štatisticky významný rozdiel vzhľadom na výsledky v postteste.

Z uvedeného je zrejmé, že izolovaná aplikácia klasickej výučby, resp. izolovaná aplikácia e-learningového modelu výučby majú približne podobné, resp. porovnateľné efekty a závery. Naopak, kombinácia uvedených režimov má za následok podstatné skvalitnenie jazykovej prípravy a zdokonalenie jazykových schopností študenta.

4. Záver

Zo skutočností uvedených vyššie vyplýva jednoznačný záver, že kombinovaná forma vzdelávania – štúdia je efektívnejšia ako len klasická forma alebo len e-learningová forma vzdelávania. Rovnako je potrebné dodať, že aj keď v niektorých oblastiach mohla byť vykazovaná osobitná úspešnosť v jednom, príp. druhom modeli, hybridný model vzdelávania sa ako celok javí najefektívnejším a najproduktívnejším pri zdokonaľovaní komunikačných, resp. jazykových schopností študenta.

5. Literatúra

- [1]ANDĚL, J.: Statistické metody. MATFYZPRESS, Praha (2003), ISBN 80-86732-08-9
- [2]KURUCOVÁ, Z.: E-learning ako prostriedok a zdroj hybridných dopadov na výchovu a vzdelávanie. IN Notitiae ex Academia Bratislavensi Iurisprudentiae, Bratislava, I/2010, str.56 a nasl.
- [3] KURUCOVÁ, Z.: Vybrané modely E-learningu vo vyspelých spoločnostiach. IN Notitiae ex Academia Bratislavensi Iurisprudentiae, Bratislava, III/2010, str.69 a nasl.
- [4]KURUCOVÁ, Z.: Úvod do tvorby efektívneho e-learningového kurzu. IN Notitiae ex Academia Bratislavensi Iurisprudentiae, Bratislava, IV/2010, str.88 a nasl.
- [5]STEHLÍKOVÁ, B. - TIRPÁKOVÁ, A. - POMĚNKOVÁ, J. - MARKECHOVÁ, D.: Metodologie výzkumu a statistická inference = Research Methodology and Statistical Inference. Brno : Mendelova Zemědělská a Lesnická Univerzita, 2009, 325 s. ISBN 978-80- 7375-362-7
- [6]TIRPÁKOVÁ, A.-MARKECHOVÁ, D.: Štatistika v praxi. FPPV UKF v Nitre (2008), 390 strán, ISBN 978-80-8094-283-0

Adresa autorov:

Zuzana Kurucová,
Fakulta práva BVŠP
Tomášiková 20
851 05 Bratislava
zuzana.kurucova@uninova.sk

Beáta Stehlíková, prof. RNDr., CSc.
Fakulta ekonomie a podnikania BVŠP
Tematínska 10
851 05 Bratislava
stehlikovab@gmail.com

Anna Tirpáková, prof. RNDr., CSc.
Katedra Matematiky UKF
Tr. A. Hlinku 1
949 01 Nitra1
atirpakova@ukf.sk

Príspevok je financovaný z projektu FCVV FPPV UKF Nitra.

Semi-parametrický prístup k odhadovaniu v autoregresnom modeli Semi-parametric approach for fitting of autoregressive model

Peter Laník

Abstract: The aim of the article is robust estimate of parameters in autoregressive model for stationary linear time series. Robust in terms of misspecified distribution of noise. Generalized moment of method is used in semi-parametric approach.

Kľúčové slová: AR(p) model, GMM, semi-parametrický model, robustný odhad.

JEL classification: C1; C5.

1. Úvod

Odfahlé pozorovania patria medzi základné problémy analýzy časových radov. Klasicky používané odhady ako metóda najšmenších štvorcov (OLS), metóda maximálnej vierohodnosti (ML), metóda podmienenej maximálnej vierohodnosti (CML), Yuleov-Walkerov odhad (YW), sú vychýlené skrz odfahlé pozorovania. Dimitris Politis v [8] navrhol algoritmus na odhad parametrov $AR(p)$ modelu. Jednoduchosť jeho myšlienky je jednou z motívácií tejto práce. Naším cieľom bude stručne zhrnúť Politisov algoritmus, popísať zovšeobecnenú momentovú metódu (GMM) pre jednorozmerné stacionárne časové rady, porovnať Politisov odhad s inými bežne používanými odhadmi ako boli spomenuté v [8]. Spracovať myšlienku Politisovho odhadu zovšeobecnenou momentovou metódou a porovnať odhady. Formulovať všeobecný semi-parametrický model pomocou GMM.

2. Metódy

Základné predpoklady:

- dáta generujúci proces $\{Y_t\}$ je slabo stacionárny,
- $EY_t = 0$, $\gamma_k = Cov(Y_t, Y_{t+k})$ pre všetky $t, k \in \mathbb{Z}$,
- $Y_t = a_1 Y_{t-1} + \dots + a_p Y_{t-p} + e_t$ pre všetky $t \in \mathbb{Z}$,
- proces $\{e_t\}$ je slabo stacionárny biely šum.

Politisov algoritmus

Z odvodenia Yuleových Walkerových (YW) rovníc je známe, že

$$\gamma_k = a_1 \gamma_{k-1} + \dots + a_p \gamma_{k-p} + \sigma^2 \delta_k, \quad k = 0, 1, \dots, p, \quad \delta_0 = 1, \delta_{k \neq 0} = 0,$$

dosadením empirických odhadov autokovariančnej funkcie v bodoch $k = 0, 1, \dots, p$ dostávame $p+1$ rovníc v tvare

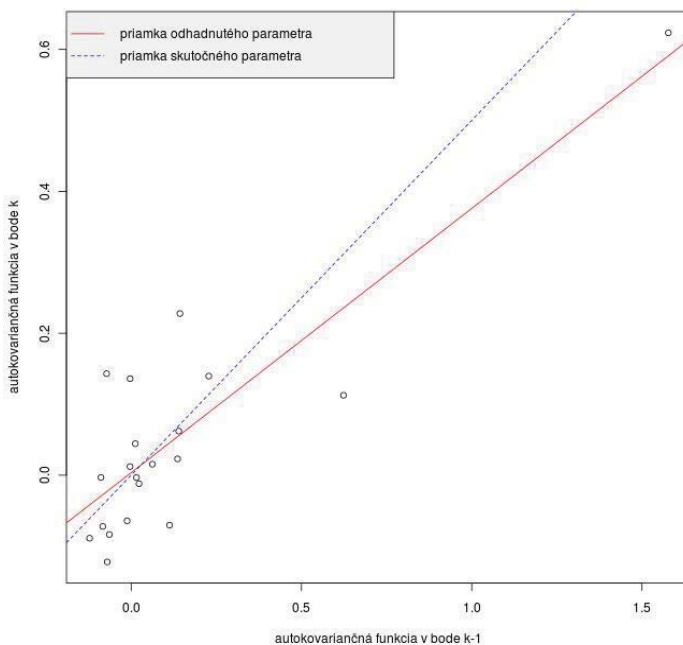
$$\hat{\gamma}_k = a_1 \hat{\gamma}_{k-1} + \dots + a_p \hat{\gamma}_{k-p} + \sigma^2 \delta_k.$$

Táto sústava má jediné riešenie $\hat{a}_1, \dots, \hat{a}_p$ a $\hat{\sigma}^2$, známe ako YW odhad. YW odhad je asymptoticky eficientný za podmienky, že dáta generujúci proces je gaussovský, podrobnosti viď [1]. Dôležitý je fakt, že na YW odhad sa dá pozerať ako na lineárnu funkciu $\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_p$. Zdalo by sa prirodzené, hľadať odhad $a_1, \dots, a_p, \sigma^2$ ako funkciu dostupnej informácie do času T na základe $\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_{T-1}$. Avšak odhad $\hat{\gamma}_k$ je kvôli vychýleniu prakticky použiteľný len pre

$k \leq T/4$. Politis navrhuje zvolit' nejaké p' tak, aby bolo väčšie ako p , ale menšie ako T , resp. požaduje, aby $n - p'$ bolo veľké, teda $p \leq p' \leq cT$ $c \in (0,1)$.

Asymptoticky nezáleží na voľbe c , pre konečné výbery Politis volí c z intervalu $[0.1, 0.2]$ a $p' = \max(p, \lfloor cT \rfloor)$.

Algoritmus odhaduje primárne parametre a_1, \dots, a_p , parameter σ^2 sa potom dopočíta z rovnice autokovariancie pre $k=0$. Pre ľahšiu predstavu o samotnom algoritme Politis uvádza jednoduchý príklad na procese $AR(1)$. YW odhad parametra a_1 je $\hat{a}_1 = \hat{\gamma}_1 / \hat{\gamma}_0$, avšak sú tu aj ďalšie alternatívne odhady $\hat{\gamma}_2 / \hat{\gamma}_1, \hat{\gamma}_3 / \hat{\gamma}_2, \dots$. Takže sa núka otázka, či je možné nejakým spôsobom skombinovať všetky tieto odhady tak, aby sme získali odhad, ktorý je najbližšie skutočnej hodnote a_1 . Politisovou odpoveďou je predstaviť si $\hat{\gamma}_k$, $k=0,1,\dots,p'$ ako pozorovania. Pre tieto pozorovania si nakresliť párový graf $\hat{\gamma}_k, \hat{\gamma}_{k-1}$. Každý bod spolu s počiatkom určuje priamku, ktorej smernica je \hat{a}_1 . Idea algoritmu (obrázok 1) je preložiť cez tieto body priamku prechádzajúcu počiatkom a jej smernicu zobrať za odhad \hat{a}_1 .



Obrázok 1 Princíp Politisovho algoritmu pre $AR(1)$.

Algoritmus:

$$\text{Označme } \underline{a}_p = (a_1, a_2, \dots, a_p)^T, \underline{\gamma}_k = (\gamma_1, \gamma_2, \dots, \gamma_k)^T \text{ a } \hat{\underline{\gamma}}_k = (\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_k)^T,$$

$$\Gamma_{p',p} = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{1-j} & \cdots & \gamma_{1-p} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{2-j} & \cdots & \gamma_{2-p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \gamma_{p'-1} & \gamma_{p'-2} & \cdots & \gamma_{p'-j} & \cdots & \gamma_{p'-p} \end{pmatrix} \mathbf{a}$$

$$\hat{\Gamma}_{p',p} = \begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{1-j} & \cdots & \hat{\gamma}_{1-p} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{2-j} & \cdots & \hat{\gamma}_{2-p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{p'-1} & \hat{\gamma}_{p'-2} & \cdots & \hat{\gamma}_{p'-j} & \cdots & \hat{\gamma}_{p'-p} \end{pmatrix} \cdot$$

Autokovariančné rovnice pre $k = 1, \dots, p'$ môžeme zapísať maticovo

$$\underline{\gamma}_{p'} = \Gamma_{p',p} \mathbf{a}_{p'}$$

pomocou empirických odhadov

$$\hat{\underline{\gamma}}_{p'} = \hat{\Gamma}_{p',p} \mathbf{a}_{p'}$$

Na poslednú rovnicu sa môžeme pozerat' ako na regresiu s odozvou $\hat{\underline{\gamma}}_{p'}$ a maticou plánu $\hat{\Gamma}_{p',p}$. Je to príklad regresie s chybami v premenných. Politis použil obvyklé riešenie pomocou OLS, čím definoval svoj odhad

$$\mathbf{a}_{p',p} = \left(\hat{\Gamma}_{p',p}^T \hat{\Gamma}_{p',p} \right)^{-1} \hat{\Gamma}_{p',p}^T \hat{\underline{\gamma}}_{p'}. \quad (1.1)$$

Špeciálnym prípadom $p' = p$ je YW odhad.

Zovšeobecnená momentová metóda

Zovšeobecnenú momentovú metódu zaviedol švédsky ekonometrik Lars Peter Hansen v [6]. Metóda je založená na minimalizácii kvadratickej formy, podobne ako *metóda minimálneho chí kvadrátu*, alebo *metóda minimálnej vzdialenosti*.

Odhadovacie rovnice, model

Nech je daný slabo stacionárny náhodný proces $\{Y(t, \omega), t \in \mathbb{T} \subset \mathbb{Z}, \omega \in \Omega\}$ definovaný na pravdepodobnostnom priestore $L^2(\Omega, \mathbf{A}, \mathbf{P})$ so stavovým priestorom \mathbf{V} . Proces je jednoznačne charakterizovaný združenou distribučnou funkciou $F_Y(y)$ a nech k tejto $F_Y(y)$ existuje združená hustota pravdepodobnosti $f_Y(y)$. Realizáciou procesu $\{Y(t, \omega)\}$ pri fixnom ω do času T je vektor $(y_T, y_{T-1}, \dots, y_1)^T = \{y_i\}_1^T$.

Definícia 1 Nech pre funkciu $u(Y_i; \theta) : \mathbf{V} \times \Theta \rightarrow \mathbb{R}^q$, kde $\theta \in \Theta \subseteq \mathbb{R}^p$, platí

1. je spojitá na Θ pre každé $Y_i \in \mathbf{V}$,
2. $E[u(Y_i; \theta)]$ existuje a je konečná pre všetky $\theta \in \Theta$,
3. $E[u(Y_i; \theta)]$ je spojitá na Θ .

Potom funkcie u nazývame odhadovacie funkcie.

Keďže uvažujeme slabo stacionárny náhodný proces, našim cieľom je odhadnúť prvé dva momenty $f_Y(y)$. Predpokladáme, že prvé momenty $f_Y(y)$ závisia od parametra $\theta \in \Theta \subseteq \mathbb{R}^p$, skutočnú hodnotu parametra označme θ_0 . Vzťah prvých dvoch momentov

$f_Y(y)$ a θ je vyjadrený cez odhadovacie rovnice $E[u(Y_i; \theta)] = 0$. Stredná hodnota odhadovacích rovníc sa nazýva teoretická momentová podmienka.

Model:

$$\Phi(\Theta) = \bigcup_{\theta \in \Theta} \Phi(\theta),$$

kde

$$\Phi(\theta) = \{f_Y(y; \theta) : E[u(Y_i; \theta)] = 0\}.$$

Ďalej uvažujme korektné špecifikovaný model, teda $\theta_0 \in \Theta$ rieši $E[u(Y_i; \theta_0)] = 0$. Aby sme spojili semi-parametrický model $\Phi(\Theta)$ s dátami $\{y_i\}_1^T$, nahradíme teoretický model $\Phi(\Theta)$ analógickým empirickým modelom

$$\Phi_T(\Theta) = \bigcup_{\theta \in \Theta} \Phi_T(\theta),$$

kde

$$\Phi_T(\theta) = \{f_Y(y; \theta) : T^{-1} \sum_{i=1}^T u(Y_i; \theta) \cong 0\}$$

a $T^{-1} \sum_{i=1}^T u(\theta, Y_i) \cong 0$ sú výberové odhadovacie rovnice. Symbol \cong znamená približne rovné. Nasledne vyberáme $\Phi_T(\theta)$ z $\Phi_T(\Theta)$ na základe nejakého kritéria, čím získame odhad.

GMM odhad

Hansenová formulácia odhadu je nasledovná: na základe pozorovaných dát chceme z q rovníc ($u(Y_i; \theta) = 0$) odhadnúť p parametrov, pričom $q > p$ teda sústava nemá (ani teoreticky) exaktné riešenie. Rovnice nahradíme odhadovacími rovnicami a zároveň dáme väčšiu váhu tým rovnicam, ktoré sú bližšie k nule.

Definícia 2 *GMM odhadom nazveme tú hodnotu θ , ktorá minimalizuje kvadratickú formu*

$$Q_T(\theta) = T^{-1} \sum_{i=1}^T u(Y_i; \theta)^T W_T T^{-1} \sum_{i=1}^T u(Y_i; \theta), \quad (1.2)$$

kde W_T je náhodná pozitívne semi-definitná matica váh, ktorá konverguje podľa pravdepodobnosti k pozitívne definitnej matici $q \times q$ konštánt.

Označme odhad $E[u(Y_i; \theta)]$ na základe informácie do času T ako $g_T(\theta) = T^{-1} \sum_{i=1}^T u(Y_i; \theta)$. Tento odhad sa často nazýva výberový (empirický) moment.

Podrobnosti o GMM možno nájsť v [4].

Simulácie

Naša simulačná štúdia vychádza zo simulačnej štúdie prezentovanej v [8]. Budeme generovať časový rad z gaussovského $AR(1)$ modelu s parametrami $a_1 = 0.5, \sigma = 1$ s počtom pozorovaní 50 a nahradíme 25-te a 26-te pozorovanie dvojicou čísiel $\{[5,5]; [5,0]; [0,5]; [0,0]; [-5,5]\}$ čím získame 6 rôznych časových radov. Pre každý časový rad odhadneme parameter $\theta = (a_1)$. Postup zopakujeme 10 000 krát, na vyhodnotenie použijeme odmocninu zo strednej kvadratickej chyby (*RMSE*).

V prípade GMM odhadu pri odhade váhovej matice, použijeme kvadraticko-spektrálny kernel a šírku kernelového okna volíme podľa Andrewsova. V prípade cue GMM nepoužívame prebielenie a zafarbenie. V prípade iteratívneho GMM použijeme prebielenie a zafarbenie s modelom $VAR(1)$ pre centrovany proces, ktorý odhadneme pomocou OLS.

3. Výsledky

Simulácie založené na empirickej autokovariancii

Použitá odhad:

- $\bar{y} = 1/T \sum_{i=1}^T y_i$
- $\tilde{\gamma}_j = 1/T \sum_{i=j+1}^T (y_i - \bar{y})(y_{i-j} - \bar{y}), \quad j = 0, 1, 2, \dots, T-1$
- $\hat{\gamma}_j = 1/(T-k) \sum_{i=k+1}^T (y_i - \bar{y})(y_{i-j} - \bar{y}), \quad k = \max(j)^1$

$$\theta = (a_1) \quad u(Y; \theta) = \begin{pmatrix} \hat{\gamma}_1 / \tilde{\gamma}_0 - \theta \\ \hat{\gamma}_2 / \tilde{\gamma}_1 - \theta \\ \vdots \\ \hat{\gamma}_k / \tilde{\gamma}_{k-1} - \theta \end{pmatrix}$$

Tabuľka 1: RMSE parametra a_1 pre rôzne odhadovacie metódy

metóda		odľahlé pozorovania					
		bez	[5,5]	[5,0]	[0,5]	[0,0]	[-5,5]
ML		0.138	0.110	0.234	0.235	0.143	0.514
CML		0.137	0.111	0.235	0.235	0.143	0.514
Politis		0.139	0.131	0.221	0.221	0.145	0.494
	2 rovnice	0.170	0.216	0.254	0.253	0.176	0.518
	3 rovnice	0.189	0.259	0.263	0.262	0.196	0.502
	4 rovnice	0.203	0.281	0.274	0.272	0.213	0.488
GMM cue	6 rovníc	0.232	0.309	0.293	0.288	0.244	0.462
	8 rovníc	0.262	0.331	0.313	0.314	0.274	0.457

Simulácie založené na podmienenom rozdelení

$$\theta = (a_1) \quad u(Y; \theta) = \begin{pmatrix} e_i \\ e_i y_{i-1} \\ e_i / (1 + e_i^2) \\ e_i y_{i-1} / (1 + e_i^2) \end{pmatrix},$$

kde $e_i = y_i - \bar{y} - \theta(y_{i-1} - \bar{y})$. Prvé dve funkcie sa vzťahujú k CML pre normálne rozdelený šum a zvyšné k CML pre šum so studentovým rozdelením s 1 stupňom voľnosti.

Tabuľka 2: RMSE parametra a_1 pre rôzne odhadovacie metódy

metóda		odľahlé pozorovania					
		bez	[5,5]	[5,0]	[0,5]	[0,0]	[-5,5]
ML		0.138	0.110	0.234	0.235	0.143	0.514
CML		0.137	0.111	0.235	0.235	0.143	0.514
Politis		0.139	0.131	0.221	0.221	0.145	0.494
GMM cue		0.168	0.174	0.203	0.199	0.173	0.246

¹ Ak chceme počítať s piatimi autokovarianciami, tak $k = 5$

4. Záver

Odhad získaný Politisovým algoritmom je z hľadiska RMSE podobný ML resp. CML odhadu. Odhad založený na spracovaní Politisovej myšlienky skrz GMM (tabuľka 1) z hľadiska RMSE horší ako Politisov odhad. Pridávanie odhadovacích rovníc znižuje RMSE v prípade scenáru [-5,5] a naopak v ostatných scenároch sa RMSE zväčšuje. Hoci sa GMM javí ako výborná alternatíva ku regresii z Politisového algoritmu, ukázali sme, že v prípade 50 pozorovaní tomu tak nieje.

Kedže sa na problém odľahlých pozorovaní dá pozeráť ako na problém zle špecifikovaného pravdepodobnostného rozdelenia šumu, použili sme ako odhadovacie rovnice skórové rovnice z CML pre gaussovský šum a šum z t rozdelenia s 1 stupňom voľnosti (tabuľka 2). RMSE je v prvých piatich scenároch podobné a pri scenári [-5,5] o polovicu menšie.

Prístup k odľahlým pozorovaniam pomocou GMM a odhadovacích rovníc založených na skórových rovniciach CML pre rôzne rozdelenia šumu sa ukazuje ako dobrá cesta a je motiváciou ďalšieho výskumu.

5. Literatúra

- [1]P.J. Brockwell and R.A. Davis. 2009. Time series: theory and methods. Springer Verlag, 2009.
- [2]P. Chaussé. 2010. Computing Generalized Method of Moments and Generalized Empirical Likelihood with R. Journal of Statistical Software, 34(11):1-35,2010.
- [3]M. Grendár and G.G. Judge. 2010. Maximum likelihood with estimating equations. 2010.
- [4]A.R. Hall. 2005. Generalized method of moments. Oxford University Press, USA,2005.
- [5]J.D. Hamilton. 1994. Time series analysis. Princeton Univ Pr, 1994.
- [6]L.P. Hansen. 1982. Large sample properties of generalized method of moments estimators. Econometrica: Journal of the Econometric Society, pages 1029-1054, 1982.
- [7]L.P. Hansen, J. Heaton, and A. Yaron. 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business & Economic Statistics,14(3):262-280, 1996.
- [8]D.N. Politis. 2009. An algorithm for robust fitting of autoregressive models. Economics Letters, 102(2):128-131, 2009.
- [9]A. Spanos. 1986. Statistical foundations of econometric modelling. Cambridge Univ Pr, 1986.

Adresa autora:

Peter Laník, Mgr.
Univerzita Mateja Bela v Banskej Bystrici
Fakulta prírodných vied
KM
Tajovského 40
974 01 Banská Bystrica
lanik@fpv.umb.sk

Výhody grafickej knižnice lattice v programe R The advantages of the package lattice in software R

Lenka Lašová

Abstract: The statistical software R is very popular in the society of statistics. It is very useful for the visualization of data. In this paper we refer some basic advantages of using this program especially its graphic package lattice. Our aim is to show the possibility of using visualization for finding the outliers in data.

Key words: program R, package lattice, trellis object, visualization of data, unusual data.

Kľúčové slová: program R, knižnica lattice, objekt trellis, vizualizácia dát, nezvyčajné pozorovania.

JEL classification: C8

1. Úvod

Knižnica lattice sa dá voľne stiahnuť z internetu ako doplnkový balíček k štatistickému programu R, ktorý je jedným z najčastejšie používaných nekomerčných štatistických softvérov. Ide o objektovo orientovaný jazyk založený na jazyku S.

V článku sa zameriame na jeho využitie pri grafickom zobrazovaní dát, ktoré si ukážeme na príkladoch. Dáta, s ktorými budeme pracovať sú voľne dostupné v programe R. Naším hlavným cieľom bude ukázať ako sme využili zobrazovanie dát pomocou lattice grafiky pri analyzovaní dát. Zameriame sa na hľadanie nezvyčajných pozorovaní v dátach.

2. Objekty triedy trellis

Mnoho funkcií v programe R nevytvára priamo výstup, ale vráti objekt, ktorý možno priradiť do premennej, prípadne použiť ako argument v ďalšej funkcii. Každý takýto objekt má svoju triedu, ktorá ho nejakým spôsobom charakterizuje a dáva nám možnosti s ním ďalej pracovať. Pokiaľ chceme objekt zobrazit' môžeme použiť funkciu *print(objekt)*, alebo pokiaľ výsledok výrazu nepriradíme do premennej, zobrazí sa výstup automaticky. Nevýhodou grafiky v základnom balíčku programu R je, že sa nám graficky výstup nedá uložiť ako objekt, len sa priamo zobrazí v grafickom okne programu. Túto možnosť však ponúka rozšírenie lattice, ktoré nám umožňuje priradovať aj grafy do premenných. Trieda takýchto objektov sa nazýva trellis. Pokiaľ máme nejaký objekt, informácie o tom z akej triedy pochádza môžeme získať pomocou príkazu *class(objekt)* alebo podrobnejšie informácie o obsahu objektu dostaneme použitím funkcie *summary(objekt)*. Samozrejme grafy nemusíme priradovať do premenných a môžeme ich používať ako v základnom balíku a ihneď zobrazit' v grafickom okne.

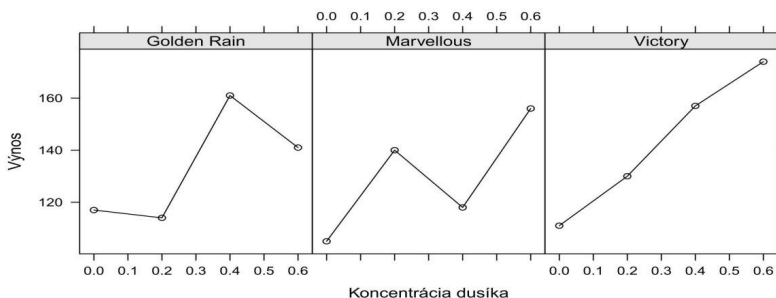
Podobne ako v základnom balíku programu R sa vstupné dáta do grafu udávajú formulou. Typická formula objektov typu trellis má nasledujúci tvar: $y \sim x | z$. Kde y je vektor hodnôt, ktoré zobrazujeme na y -ovu os, x je vektor hodnôt, ktoré zobrazujeme na x -ovu os a z je triediaca premenná. Grafický výstup sa bude skladať z panelových grafov, ktorých budeme mať toľko, koľko je rôznych hodnôt v rámci vektora z a v každom takomto grafe budú zobrazené namerané hodnoty y voči hodnotám vektora x , ktoré nadobúdajú zodpovedajúcu hodnotu vektoru z . Teda oproti základnému balíku tu môžeme pridať triediaci vektor z .

Teraz si využitie triediacej premennej ukážeme na jednoduchom príklade s využitím grafu *xyplot(formula, data)*. Dáta, ktoré budeme zobrazovať sú zadané v tabuľke a je možné ich stiahnuť prostredníctvom programu R.

Tabuľka 1: Dáta Ovovs

Koncentrácia dusíka	Výnos	Odroda ovsa	Umiestnenie
0.0	111	Victory	1
0.2	130	Victory	1
0.4	157	Victory	2
0.6	174	Victory	2
0.0	117	Golden Rain	1
0.2	114	Golden Rain	1
0.4	161	Golden Rain	2
0.6	141	Golden Rain	2
0.0	105	Marvellous	1
0.2	140	Marvellous	1
0.4	118	Marvellous	2
0.6	156	Marvellous	2

Budeme zobrazovať výnos voči koncentrácii dusíka pomocou spojnicového grafu a ako triediacu premennú sme si zvolili odrodu ovsa.



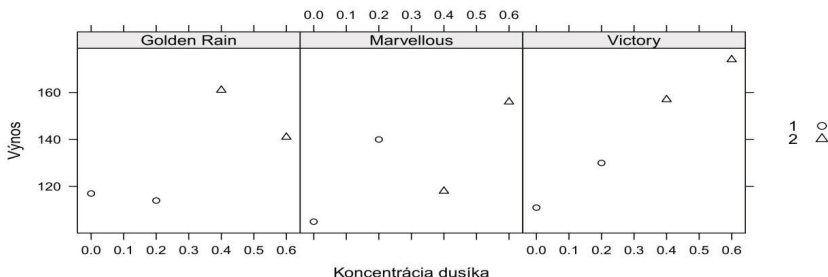
Obrázok 1: Zobrazenie dát o výnose ovsa

Poznamenajme, že sa v praxi stretávame aj s prípadom, že chceme použiť ako triediacu premennú spojenú veličinu. V tomto prípade môžeme premennú roztriediť do intervalov a s nimi pracovať ako s kategóriami. Na tento účel je možné v programe R využiť dve funkcie. Príkazom `cut(data,n)` rozdelíme premennú na n rovnako veľkých disjunkčných intervalov. Pomocou príkazu `equal.count(data,n)` dostaneme n intervalov, ktoré majú približne rovnaké početnosti, intervaly sa môžu prekrývať.

Keďže balíček `lattice` nám dáva možnosť pracovať s grafom ako s objektom, uvedieme si niekoľko výhod takejto voľby. Ako už bolo spomínané graf, ktorý sme vytvorili a priradili do premennej `graf`, si môžeme kedykoľvek opäť vyvolať funkciu `print(graf)`. Graf môžeme vnímať ako mriežku so stĺpcami a riadkami, a jednotlivé vnútorné grafy budeme nazývať panely resp. panelové grafy. Pomocou funkcie `print` môžeme vykresliť nielen celý graf ale aj jednotlivé panelové grafy napr. príkaz `print(graf[2])` nám vykreslí len ten graf, ktorý zahŕňa dáta s hodnotou triediacej premennej `Marvellous`. Pomocou funkcie `dim(graf)` vieme zistiť koľko kategórií má triediaca premenná a príkaz `dimnames(graf)` nám vypíše aj názvy

jednotlivých kategórii tejto premennej. Nespornou výhodou je aj možnosť použitia funkcie *update(graf,...)* pomocou ktorej môžeme meniť nami vybrané parametre v už vytvorenom grafe.

Ďalšou dôležitou výhodou balíčka *lattice* je možnosť pridať do grafu ďalšiu premennú pomocou parametra *group*. Použijeme dáta z tabuľky 1., tento krát však nebudeme spájať namerané hodnoty čiarou a pridáme vektor Umiestnenie ako rozlišujúcu premennú.



Obrázok 2: Zobrazenie dát o výnose ovsa s rozlišujúcou premennou

3. Niektoré grafické parametre

Čo sa týka voľby grafických parametrov, nie je možné spomenúť všetky, preto sme vybrali len tie najpoužívanjšie. Jedným z nich je napríklad parameter *aspect* pomocou, ktorého môžeme meniť pomer strán y/x na jednotlivých panelových grafoch, tento parameter je prednastavený tak aby celý graf, čo najviac vyplňal grafické okno. Pokiaľ by sme chceli medzi jednotlivé grafy vložiť medzery môžeme tak urobiť pomocou parametra *between*. Parametrom *layout* určujeme rozmery výsledného grafu v zmysle počtu stĺpcov a riadkov. Čo sa týka farieb, je možné ich nastaviť pomocou parametrov *col.line* a *col.symbol* na voľbu farby čiary a symbolu. Jedným z najčastejšie používaných parametrov je určite parameter *type*, ktorý je dostupný aj v základnom balíčku programu R, ale v knižnici *lattice* je rozšírený o niektoré užitočné možnosti voľby tohto parametra. Voľbou tohto parametru zadávame o aký typ grafu máme záujem. Medzi základné preddefinované možnosti patrí klasický bodový a spojnicový graf, prípadne zobrazenie bodov pospájaných čiarami, zobrazenie hodnôt pomocou vertikálnych čiar alebo schodový graf. V knižnici *lattice* máme navyše tieto varianty: dáta zobraziť ako body a pridať k nim lineárnu regresnú krivku, pridať odhad regresnej funkcie pomocou lokálnej polynomickej regresie, pridať spriemerovanú čiaru, prípadne zobraziť mriežku v grafe. Samozrejme máme aj parametre, ktorými môžeme pridávať popisy do grafu napr. *xlab*, *ylob* doplní popis k osiam, *main* doplní názov grafu nad graf a pomocou *sub* môžeme napísať popis pod graf.

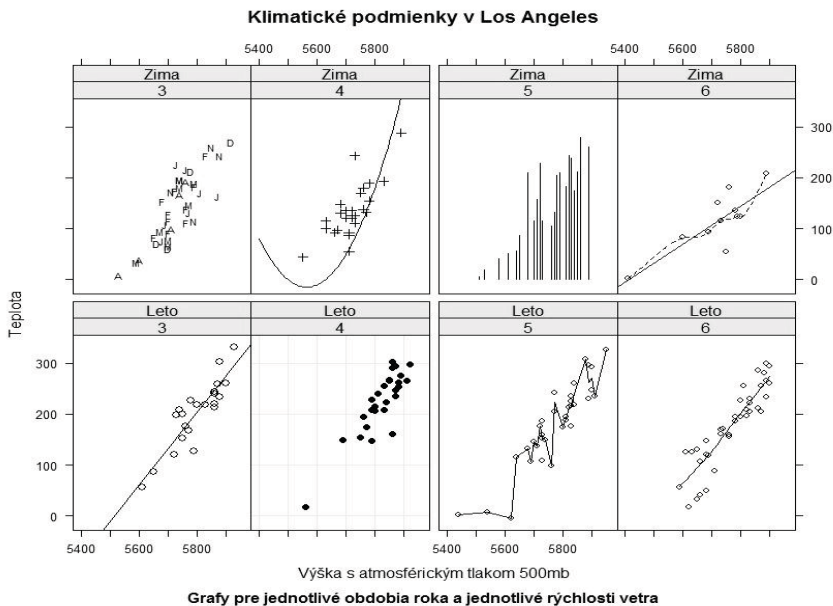
4. Panelové funkcie

V predchádzajúcej časti sme popisali aké grafické parametre môžeme použiť pri vykresľovaní grafu. Samozrejme je prirodzenou požiadavkou, aby sme si mohli zvoliť pre jednotlivé panelové grafy rôzne nastavenia. Knižnica *lattice* nám to umožňuje prostredníctvom funkcie *panel*, ktorú si môžeme vložiť ako jeden z parametrov grafu. Na nasledujúcom obrázku si ukážeme aký výstup môžeme dostať.

Pracovali sme s dátami dostupnými prostredníctvom programu R v knižnici *faraway*. Dáta obsahujú namerané niektoré ukazovatele počasia počas dní jedného roku v Los Angeles. My sme využili nameranú teplotu a nameranú výšku v metroch, v ktorej je výška

atmosférického tlaku 500mb. Dni sme použili pri vytváraní dvoch premenných. Prvá obsahuje údaje o mesiaci, v ktorom boli namerané hodnoty a druhá obsahuje ročné obdobie, do ktorého prislúcha daný mesiac. Využili sme aj vektor obsahujúci nameranú silu vetru v daný deň. Tieto dáta sme použili pri vytvorení obrázku 3, aby sme ukázali, aké možnosti nám ponúka knižnica lattice. Ako môžeme vidieť zobrazovali sme závislosť teploty od výšky s atmosférickým tlakom 500mb. Ako triediacu premennú sme použili dve premenné a to ročné obdobie s kategóriami Zima a Leto, a druhou použitou premennou bola sila vetra, z ktorej sme chceli zobrazíť len kategórie označené 3,4,5,6. Takže pre každú kombináciu kategórii z prvej a druhej triediacej premennej máme samostatný graf. Pomocou príkazu *between* sme oddelili grafy, ktoré obsahujú dáta zo Zimy a z Leta. Takisto sme oddelili grafy obsahujúce dáta z dní, kedy bola sila vetra 3,4 a dní kedy bola sila vetra 5,6. Využili možnosť obrázok popísať: pridalí sme názvy osí, názov graf aj popis pod graf.

Snažili sme sa ukázať viaceré možnosti využitia funkcií knižnice lattice, preto sme pre jednotlivé panelové grafy použili iné hodnoty parametra *type*, menili sme aj možnosti zobrazenia symbolov. V grafoch Leto/3 a Zima/6 sme k zobrazeným bodom pridalí aj lineárny odhad a na porovnanie v grafe Zima/6 máme aj odhad regresnej funkcie pomocou lokálnej polynomickej regresie, ktorý sme pridalí aj do grafu Leto/6. V grafe Leto/5 sme k dátam vložili priemerovú funkciu a do grafu Zima/4 sme vložili nami zadanú krivku. Menili sme aj tvary jednotlivých symbolov napr. v Zima/4 sme použili krížik, v Leto/4 sme dané symboly vyplnili a v Leto/3 sme zvolili väčšiu veľkosť. Špeciálne symboly sme zvolili v grafe Zima/3 kde nám každý symbol reprezentuje mesiac, v ktorom bolo meranie spravené. V grafe Zima/5 sme zvolili iné zakreslenie dát pomocou vertikálnych čiar. V grafe Leto/4 sme pomocou *type* pridalí aj mriežku.

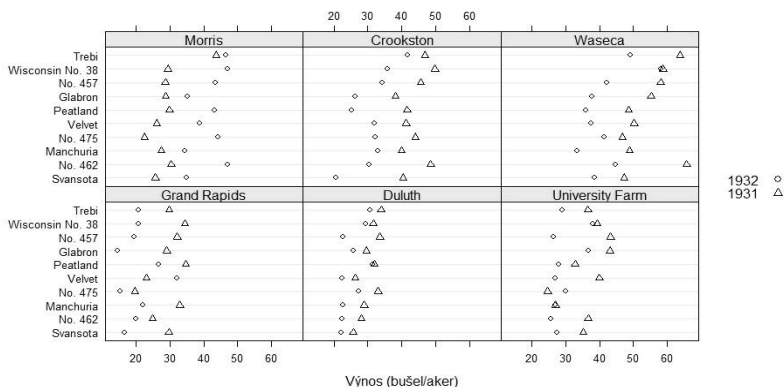


Obrázok 3: Zobrazenie dát o klimatických podmienkach v Los Angeles

5. Využitie lattice grafov pri hľadani nezvyčajných pozorovaní

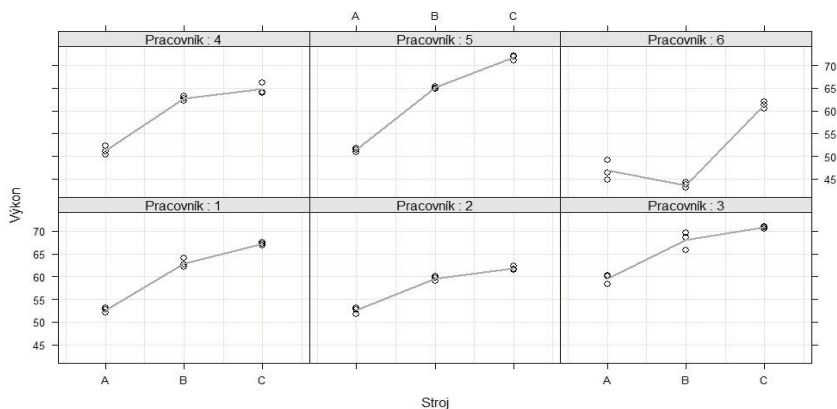
Vhodné zobrazenie skúmaných dát nám často krát pomáha s ich analýzou. Jedným z dôležitých cieľov zobrazovania dát je identifikácia nezvyčajných pozorovaní. S touto úlohou nám práve knižnica lattice môže pomôcť. Uvedieme si názorný príklad, v ktorom nám vhodné zobrazenie jednoducho odhalí prípad nezvyčajného pozorovania v dátach.

Dáta obsahujú výnosy jednotlivých odrôd jačmeňa na šiestich experimentálnych miestach v Minnesote za roky 1931 a 1932. Ako môžeme vidieť na obrázku 4, miesto Morris sa od všetkých ostatných líši tým, že všetky miesta mali výnosy jednotlivých odrôd vyššie v roku 1931 ako v 1932, ale Morris má výšku výnosov presne naopak.



Obrázok 4: Zobrazenie dát o výnose jačmeňa

Zobrazenie pomocou knižnice lattice môžeme využiť aj pri analýze dát o výkonnosti pracovníkov. Všimnime si na obrázku 5, že priemerová funkcia je u všetkých pracovníkov približne rovnaká, ale šiesty pracovník sa odlišuje od ostatných.



Obrázok 5: Zobrazenie dát o výkonnosti pracovníkov

6. Záver

Pri analyzovaní dát je nevyhnutné namerané dáta vhodne zobrazovať, na to potrebujeme efektívne nástroje. Za najväčšie výhody knižnice lattice považujeme možnosť jednoducho vytvoriť grafy s triediacou a rozlišujúcou premennou. Nespornou výhodou je, že program R aj s doplnkovým balíčkom lattice je voľne dostupný na internete. Výhodou pre pokročilých užívateľov ovládajúcich jazyk S / R je možnosť spraviť akýkoľvek graf bez obmedzení. V práci sme sa snažili ukázať, že takéto zobrazovanie je vhodné napríklad pri objavovaní nezvyčajných pozorovaní v dátach. Samozrejme pomocou lattice zobrazenia môžeme pozorovať aj rozdiely medzi jednotlivými skupinami pozorovaní. Vďaka tomu môžeme napríklad zistiť, kedy je vhodnejšie niektoré dáta modelovať samostatne po skupinách (napr. často krát osobitne mužov a osobitne ženy).

7. Literatúra

- [1]BATES, D. M.. – PINHEIRO, J. C. 2000. Mixed-Effects models in S and S-plus. New York: Springer, 2000. 526 s. ISBN 0-387-98957-9.
- [2]FARAWAY, J. J. 2006. Extending the Linear Model with R. Chapman & Hall, 2006. 332 s. ISBN 0-203-62105-0.
- [3]SARKAR, D. 2008. Lattice: Multivariate Data Visualization with R. New York: Springer, 2008. 280 s. ISBN 978-0-387-75968-5.
- [4]<http://www.r-project.org/>

Tento príspevok vznikol vďaka podpore UGA pre projekt I-10-005-07.

Adresa autora:

Lenka Lašová, RNDr. PhD.
Inštitút matematiky a inf. ÚVV UMB
Tajovského 40
974 01 Banská Bystrica
Lenka.Lasova@umb.sk

Názory študentov LF UK a FaF UK na fajčenie Opinion of students LF UK and FaF UK on smoking

Erika Macháčová, Ján Luha

Abstract: In this article we present principal results from survey about opinion young people, university students, on some questions concerning smoking. We present answers from university students Faculty of Medicine and Faculty of Pharmacy on questions about first experimentation with smoking and age of first use of cigarette.

Key words: opinion, university students, smoking.

Kľúčové slová: názory, univerzitní študenti, fajčenie.

JEL Classification: C1, C12, C14, J10, J11.

1. Úvod

Fajčenie je významným rizikovým faktorom, prispievajúcim k viac ako 20 rôznym ochoreniam a predčasným úmrtiam. Cigareta je jediným spotrebným výrobkom, ktorý ak je používaný, k čomu je určený, zabíja svojho konzumenta. Užívanie tabaku celosvetovo zodpovedá za 8,8% úmrtí a podieľa sa 4,1% na predčasnej strate zdravia populácie (merané prostredníctvom indexu DALYs) [3]. Na Slovensku je situácia podstatne horšia. Fajčenie je známym a potvrdeným rizikovým faktorom kardiovaskulárnych ochorení (KVCH) mužov i žien. Výskyt a mortalita na KVCH je vyššia v ekonomicky slabších skupinách obyvateľstva.

Podobný prípad, ako je táto naprogramovaná cesta smrti a deštrukcie, je nemožné nájsť v histórii [8]. Fajčenie predstavuje však aj zdravotne závažné rizikové chovanie. Fajčiari ohrozujú nielen seba, ale expozícií tabakovému dymu vystavujú aj nefajčiarov. Títo nemajú žiadnu tzv. adaptačnú schopnosť na nežiaduce vplyvy fajčenia, ich kardiovaskulárny systém je na ne extrémne citlivý a nemôže sa prispôsobiť toxickým látkam obsiahnutým v cigaretovom dyme [1, 2].

Tolerovanie fajčenia spoločnosťou svedčí o nízkej hodnote zdravia v systéme hodnôt. Jednou z hlavných príčin podceňovania negatívnych vplyvov fajčenia na zdravie je pomerne dlhé obdobie medzi začiatkom fajčenia a prejavom sa jeho následkov, ktoré sú oveľa vyššie u tých osôb, ktoré začali fajčiť v mladom veku, v čase vývoja organizmu [8]. So zvyšovaním vzdelania preukazne klesá spotreba cigariet. Je však alarmujúce, že prvé skúsenosti s fajčením sa posúvajú do stále nižších vekových skupín. Preventívne opatrenia musia byť však zamerané už na deti a mládež, pretože prvé skúsenosti s fajčením spadajú už do obdobia základnej školskej dochádzky.

Pre znižovanie a elimináciu zdravotných rizík vyplývajúcich z fajčenia je dôležitý postoj lekárov a zdravotníckych pracovníkov k fajčeniu, ktorí musia byť príkladom a nefajčiť.

V príspevku uvádzame výsledky vlastného prieskumu názorov mladých ľudí, univerzitných študentov, ktorí odpovedali na 43 otázok o fajčení. Výskum sme realizovali pomocou štandardizovaného dotazníka zameraného na študentov medicíny, ošetrovateľstva, stomatológie a farmácie používaného v projekte „GLOBAL HEALTH PROFESSIONAL STUDENTS SURVEY“ (naša účasť na projekte pozostávala so zbieraním podkladov tretích ročníkov).

2. Základné charakteristiky prieskumu

Výskum sa uskutočnil v období od novembra 2009 do decembra 2010.

Do prieskumu bolo zaradených 1215 študentov, a to 981 z Lekárskej fakulty Univerzity Komenského v Bratislave (LF UK) a 234 študentov z Farmaceutickej fakulty UK Bratislava (Fa F UK).

Z 981 študentov LF UK bolo 930 poslucháčov všeobecného lekárstva a 51 študentov zubného lekárstva. Z FaF UK - 234 študentov boli všetci z tretieho ročníka programu farmácia.

Z 981 študentov LF UK zaradených do prieskumu študovalo v prvom ročníku 345 študentov, v treťom 309 a v piatom ročníku 327 poslucháčov všeobecného lekárstva. Z 51 študentov zubného lekárstva bolo 24 z prvého ročníka a 27 z tretieho ročníka (tab. 1.)

Tab. 1. Rozloženie výberového súboru podľa formy štúdia a ročníka štúdia

ročník štúdia	Všeobecné lekárstvo	Zubné lekárstvo	Farmácia	Total
prvý	321	24	0	345
tretí	282	27	234	543
piaty	327	0	0	327
Total	930	51	0	1215

Rozloženie súboru podľa pohlavia a formy štúdia – z 1209 študentov vykazujú prevahu ženy (tab.2.) v pomere 2,64:1 (877:332). 6 študenti neuviedli v dotazníku pohlavie.

Tab. 2. Rozloženie výberového súboru podľa formy štúdia a pohlavia

forma štúdia	Žena	Muž	Total
Všeobecné lekárstvo	651	279	930
Zubné lekárstvo	36	15	51
Farmácia	190	38	228
Total	877	332	1209

Z 1215 poslucháčov, ktorí sa zúčastnili na prieskume bol najvyšší počet vo vekovej kategórii 19 až 24 ročných – 1049 študentov (tab.3).

Tab. 3. Rozloženie výberového súboru podľa formy štúdia a vekových kategórií

forma štúdia	18 a menej	19-24	25-29	30+	Total
Všeobecné lekárstvo	39	786	100	5	930
Zubné lekárstvo	4	44	3	0	51
Farmácia	3	219	10	2	234
Total	46	1049	113	7	1215

Dotazník, okrem demografických znakov (fakulta, forma štúdia, ročník štúdia, pohlavie a vek), obsahoval 40 meritórnych otázok zameraných na niektoré problémy ohľadne fajčenia.

V príspevku podrobnejšie uvádzame výsledky odpovedí na otázky zamerané či už niekedy experimentovali s fajčením a v akom veku prvýkrát skúsili cigaretu.

Z celého súboru 1215 študentov, poslucháčov Lekárskej fakulty a Farmaceutickej fakulty, až 78,21 % experimentovalo s fajčením. Prvú skúsenosť s cigaretou do 15 rokov uviedlo až 46,% a do 17 rokov 65,7% študentov. Naše zistenia sú podobné, ako prof. Redhamera (10), ktorý uvádza, že väčšina pravidelných fajčiarov začína fajčiť pred svojim 20 rokom života, najčastejšie vo veku 13 -16 rokov.

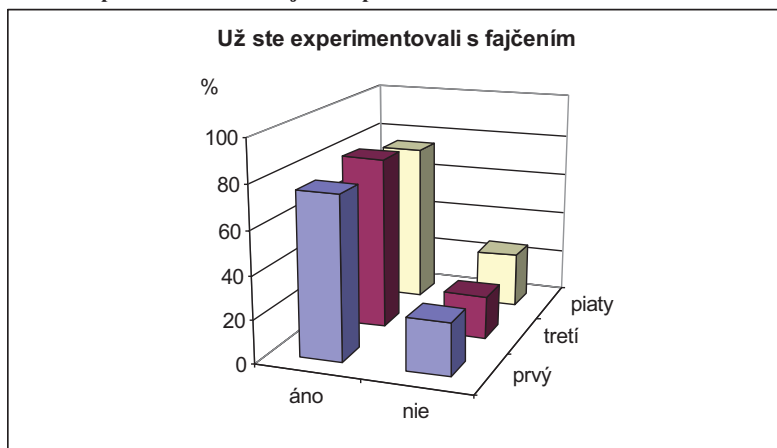
3. Komparácia názorov študentov všeobecného lekárstva LF UK podľa ročníkov 1., 3. a 5.

Zo súboru poslucháčov Lekárskej fakulty študenti 1. ročníka experimentovali s fajčením v 75,7%, 3. ročníka v 80,1% a 5. ročníka v 74,3%. Výsledky sú v kontingenčnej tabuľke 4 a znázornené na Obr.1. Rozdiely nie sú štatisticky významné i keď vidno, že študenti tretieho ročníka mali väčší podiel tých čo mali skúsenosti s fajčením, P-hodnota Fisherovho exaktného testu $P=0,211$.

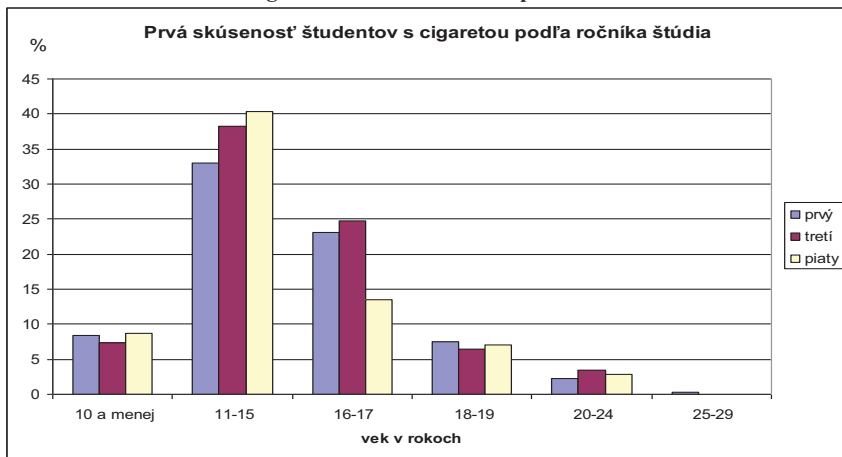
Tab. 4. Experimentovanie s fajčením študentov všeobecného lekárstva LF UK

			o1 Už ste experimentovali niekedy s fajčením cigariet,hočí len jeden, či dva ľahy?		Spolu
			1 Áno	2 Nie	
Ročník štúdia	prvý	počet	243	78	321
		% podľa ročníka štúdia	75,7%	24,3%	100,0%
	tretí	počet	226	56	282
		% podľa ročníka štúdia	80,1%	19,9%	100,0%
	piaty	počet	243	84	327
		% podľa ročníka štúdia	74,3%	25,7%	100,0%
Spolu		počet	712	218	930
		%	76,6%	23,4%	100,0%

Obr. 1. Experimentovanie s fajčením podľa ročníka štúdia študentov LF UK



Prvú skúsenosť s cigaretou študentov podľa jednotlivých ročníkov vidíme na obrázku č. 2. Najviac študentov prvýkrát vyskúšalo cigaretu vo veku do 15 rokov a tiež do 17 rokov. I keď nižšie percento, ale vo všetkých ročníkoch sme zaznamenali prvú skúsenosť s cigaretou vo veľmi nízkom veku – do 10 rokov.

Obr. 2. Prvá skúsenosť s cigaretou študentov LF UK podľa ročníka štúdia

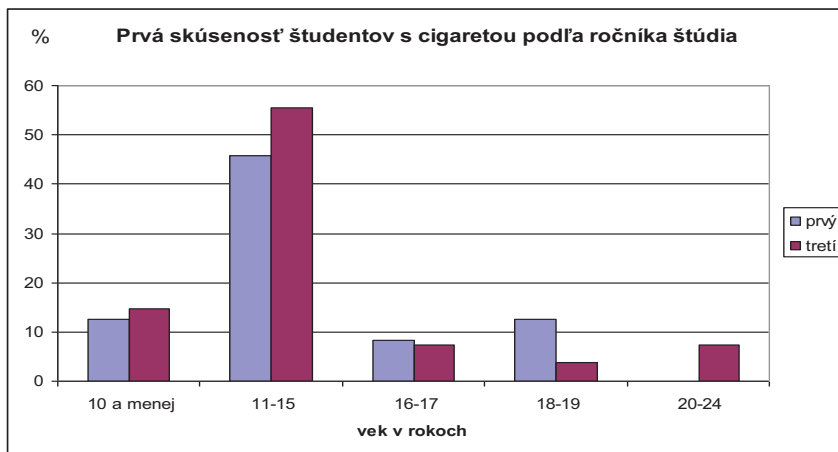
4. Komparácia názorov študentov zubného lekárstva LF UK podľa ročníkov 1. a 3.

V sledovanom súbore bolo 51 študentov zubného lekárstva v prvom a treťom ročníku štúdia. Komparácia výsledkov odpovedí na otázku či už experimentovali niekedy s fajčením je uvedená v kontingenčnej tabuľke 5. Rozdiely nie sú štatisticky významné, P-hodnota Fisherovho exaktného testu je $P=0,451$. Skúsenosť s fajčením uviedlo 88,9% študentov tretieho a 79,2% študentov prvého ročníka

Tab. 5 Experimentovanie s fajčením študentov zubného lekárstva LF UK

		o1. Už ste experimentovali niekedy s fajčením cigariet, hoci len jeden, či dva ľahy?		spolu	
		1 Áno	2 Nie		
ročník štúdia	prvý	počet	19	5	24
		% podľa ročníka štúdia	79,2%	20,8%	100,0%
	tretí	počet	24	3	27
		% podľa ročníka štúdia	88,9%	11,1%	100,0%
spolu		počet	43	8	51
		%	84,3%	15,7%	100,0%

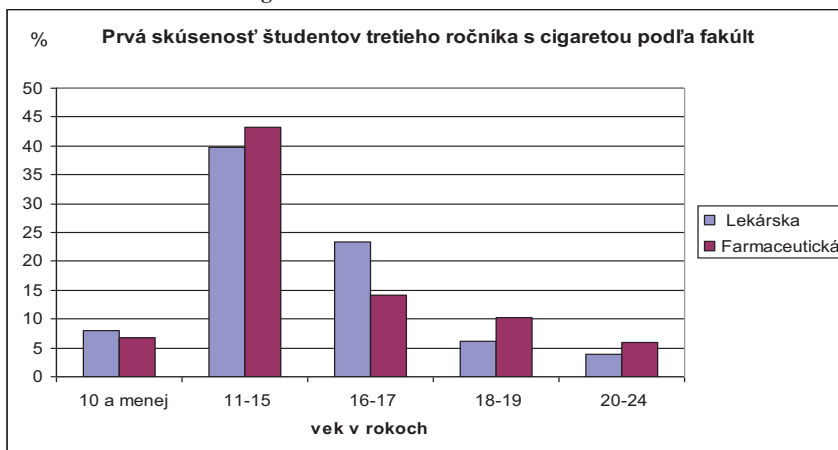
Porovnanie veku prvej skúsenosti s fajčením študentov prvého a tretieho ročníka zubného lekárstva nie je významne odlišné. P-hodnota Fisherovho exaktného testu je $P=0,629$. Výsledok prezentujeme na obr. 3.

Obr. 3. Prvá skúsenosť s fajčením študentov 1. a 3.ročníka zubného lekárstva

5. Komparácia názorov študentov 3. ročníka LF UK a študentov 3. ročníka FaF UK

Ak porovnáme tretie ročníky poslucháčov medicíny a farmaceutickej fakulty, na otázku experimentovania s fajčením cigariet odpovedalo kladne 80,9% študentov 3. ročníka LF UK (z 309 študentov) a 80,8 % študentov 3. ročníka FaF UK (z 234 študentov). Rozdiel nie je štatisticky signifikantný, P-hodnota Fisherovho exaktného testu je $P=1,000$.

Prvá skúsenosť s cigaretou do 15. rokov študenti 3. ročníka LF UK uviedli v 47,9 % a do 17 rokov v 71,2 % . 50% poslucháčov 3. ročníka FaF UK malo prvú skúsenosť s cigaretou do 15 rokov a 64,1% do 17 rokov . P-hodnota Fisherovho exaktného testu je $P=0,058$, čo je na hranici štatistickej signifikantnosti.

Obr. 4. Prvá skúsenosť s cigaretou študentov 3. ročníkov LF UK a FaF UK

6. Záver

Naše výsledky potvrdzujú zistenia mnohých autorov, že prvé skúsenosti s fajčením sa posúvajú do nižších vekových skupín a spadajú do obdobia základnej školskej dochádzky. (väčšina fajčiarov začína fajčiť už v období dospievania).

Pre znižovanie a elimináciu zdravotných rizík vyplývajúcich z fajčenia je dôležitý postoj lekárov a zdravotníckych pracovníkov k fajčeniu, ktorí by mali byť pozitívnym príkladom v systematickej primárnej prevencii detí a dospelých.

7. Literatúra

- [1] Kozák J.T (2000): Mezinárodní konference SZO o tabáku a zdraví, Čas. Lék. čes., **139**, 5, pp.158, 2000.
- [2] Glantz S. A., Parmley W. W. (1995): Passive Smoking and Heart Disease. Mechanisms and Risk, *Jama*, **273**, pp.1047-1053, 1995.
- [3] Baška T. (2008): Prevencia užívania tabaku u školskej mládeže. Národná koalícia pre kontrolu tabaku v SR v spolupráci s Jesseniovou lekárskou fakultou v Martine, UK v Bratislave. 1. vydanie, Martin 2008. 64 s. ISBN: 978-80-969767-1-3.
- [4] Kanji G. K. (2006): 100 Statistical Tests. 3rd Edition. SAGE 2006.
- [5] Luha, J. (1985): Testovanie štatistických hypotéz pri analýze súborov charakterizovaných kvalitatívnymi znakmi. STV Bratislava 1985.
- [6] Luha J. (2009): Matematicko-štatistické aspekty spracovania dotazníkových výskumov. FORUM STATISTICUM SLOVACUM 3/2009. SŠDS Bratislava 2009. ISSN 1336-7420.
- [7] Luha J. (2010): Metodologické zásady záznamu dát z rozličných oblastí medicíny a zásady ich kontroly. FORUM STATISTICUM SLOVACUM 1/2010. SŠDS Bratislava 2010. ISSN 1336-7420.
- [8] Macháčová E., Ševčíková E., Baráková A. (2000): Fajčenie ako jeden z rizikových faktorov kardiovaskulárnych ochorení. In: Profesionalita & Progres & Podpora zdravia. Fragment z dejín LFUK, Bratislava, E & J s.r.o. 2000, s. 121-124. ISBN 80-968419-1-2.
- [9] Redhamer R.: Fajčenie a zdravie. Slovak Academic Press, spol. s r.o., Bratislava, 2007, 158 s. ISBN 978-80-8095-017-0
- [10] Riffenburg R. H.: Statistics in Medicine, Second Edition. Academic Press 2005.
- [11] Ševčíková E., Argalášová E., Liu YH., Jurkovičová J., Štefaniková Z., Macháčová E., Rosinský J., Gregušová A., Karpatová E., Kuruc Š., Majzonová M., Zajacová M., Salamonová M., Šimko G., Zenka M., Weitzman M. : Ohrozenie zdravia školských detí vystavených pasívnemu fajčeniu.

Adresa autorov:

Erika Macháčová MUDr., PhD.
Ústav epidemiológie LF UK, Bratislava
erika.machacova@fmed.uniba.sk

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

Práca bola podporená grantom VEGA 1/0135/09.

**Štatistické vyhodnotenie výsledkov výskumu v hudbe
Friedmanovým testom**
**Statistical evaluation of results of the research in music
by Friedman's test**

Dagmar Markechová

Abstract: The aim of this article is to present the applications of nonparametrical statistical methods, especially the Friedman test in evaluating research project.

Key words: nonparametrical methods of mathematical statistics, Friedman's test, programme STATISTICA

Kľúčové slová: neparametrické metódy matematickej štatistiky, Friedmanov test, program STATISTICA

JEL classification: C14

1. Úvod

V rokoch 2007-2010 bol autorkou dizertačnej práce [2] realizovaný výskum, ktorého cieľom bolo okrem iného zistiť, v akej miere sú zastúpené klavírne diela slovenských hudobných skladateľov na vybraných hudobných podujatiach u nás za posledných desať rokov a či sa tieto hudobné podujatia líšia vzhľadom na výber diel slovenských hudobných skladateľov. Cieľom predloženého článku je prezentovať vyhodnotenie výsledkov tohto výskumu pomocou neparametrických metód matematickej štatistiky a ich interpretácia.

2. Materiál a metódy

V rámci výskumu, ktorý bol realizovaný v rokoch 2007-2010 autorkou dizertačnej práce [2], bol sledovaný výber klavírných diel slovenských hudobných skladateľov na vybraných hudobných podujatiach organizovaných na Slovensku. Boli to nasledujúce hudobné podujatia:

- 1) Medzinárodná súťaž Košice (MSKE)
- 2) Festival Ivana Ballu v Dolnom Kubíne (FIB)
- 3) Celoslovenská súťaž Schneidera Trnavského (CSST)
- 4) Celoslovenská súťaž Mladí klaviristi v Bratislave (CSMK)
- 5) Detský hudobný festival Jána Cikkeru v Banskej Bystrici (DHF JC)
- 6) Celoslovenská súťaž Nitrianska Lutna (CSNL)
- 7) Klavírna súťaž Dezidera Kardoša (KSDK), od roku 2010 sa súťaž koná pod názvom Klavír 20. a 21. storočia
- 8) Celoslovenská súťaž „Prehliadka slovenskej hudobnej tvorby“ s medzinárodnou účasťou v sólovej a komornej hre v Rajci (SP SHT), od roku 2010 sa súťaž koná pod názvom „Rajecká jar“
- 9) Festival súčasného umenia Košice (FSU KE)
- 10) Festival Nová slovenská hudba v Bratislave (FNŠH).

V zátvorkách sú uvedené skratky názvov podujatí, ktoré budeme v ďalšom používať. Predmetom výskumu bolo zastúpenie klavírných diel 39 slovenských hudobných skladateľov na týchto podujatiach, z ktorých uvádzame tieto známe mená: Miroslav Bázlik, Ján Cikker, Eugen Suchoň, Dezider Kardoš, Ján Zimmer, Juraj Hatrík, Andrej Očenáš, Mikuláš Schneider Trnavský, Igor Dibák, Tibor Frešo, Ilja Zeljenka, Iris Szeghy, Dušan Martinček, Július Kowalski, Ivan Parík, Vítázoslav Kubička, Ladislav Holoubek, Viliam Fíguš Bystrý, Hanuš Domanský.

Jednou z najpoužívanejších metód vo výskumnej praxi ako v prírodovedných tak aj v spoločenskovedných disciplínach je analýza rozptylu. Je to parametrická metóda, pomocou ktorej porovnáваме stredné hodnoty niekoľkých základných súborov. Najčastejším predpokladom obvykle používaných parametrických štatistických metód (ako je napríklad t -test, analýza rozptylu, intervalové odhady a testy v regresnej analýze) je normálne rozdelenie základného súboru. Ak by sme použili pri štatistickom vyhodnocovaní výsledkov výskumu parametrický test a neboli by splnené predpoklady jeho použitia (napríklad predpoklad o normálnom rozdelení základného súboru), závery získané na základe tejto metódy by boli nevierohodné. V prípade, že nie sú splnené predpoklady použitia parametrickej testovacej metódy, je možné použiť niektorú z neparametrických metód. Pri neparametrických metódach sú predpoklady pre ich použitie oveľa slabšie ako pri metódach parametrických a často ich môžeme použiť pre akékoľvek rozdelenie základného súboru. Nevýhodou neparametrických techník je však tá skutočnosť, že sú menej citlivé a presné ako metódy parametrické. Platí preto zásada, že dávame prednosť parametrickým testom, samozrejme len vtedy, keď sú splnené všetky predpoklady pre ich použitie.

Analýzu rozptylu môžeme použiť len vtedy, ak sú splnené nasledujúce podmienky:

- výberové súbory pochádzajú zo základných súborov s normálnym rozdelením;
- rozptyly základných súborov sa rovnajú;
- súbory sú navzájom nezávislé.

V prípade, že uvedené podmienky nie sú splnené a súbory sú navzájom nezávislé, pre porovnanie súborov môžeme použiť neparametrickú alternatívu analýzy rozptylu, ktorou je Kruskalov-Wallisov test. Na porovnávanie niekoľkých základných súborov na základe závislých výberových súborov, pričom nie je splnený predpoklad o normálnom rozdelení základných súborov, je vhodnou metódou Friedmanov test. Friedmanov test je zovšeobecnením Wilcoxonovho jednovýberového testu a je neparametrickou analógiou analýzy rozptylu dvojného triedenia s jedným pozorovaním v každej podtriede. Popíšeme túto metódu.

Nech X_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, k$) sú náhodné premenné so spojitými distribučnými funkciami. Chceme testovať nulovú hypotézu H_0 , že distribučné funkcie premenných $X_{i1}, X_{i2}, \dots, X_{ik}$ sú totožné. V praxi ide o tzv. *model náhodných blokov*. Skúma sa vplyv k ošetrení, ktoré sú aplikované na n blokoch. Treba rozhodnúť, či ošetrenia prinášajú rovnaký efekt. Ak nemajú rovnaký efekt, treba ďalej určiť, ktoré ošetrenia sa od seba líšia. Často nejde samozrejme o ošetrenia v pravom slova zmysle, ale jednoducho je nejaká veličina zaznamenávaná na každom sledovanom objekte (resp. bloku) v k prípadoch. Za tieto prípady budeme považovať jednotlivé hudobné podujatia, takže $k = 10$. Pozorovanou náhodnou premennou X je počet diel hudobného skladateľa zaradených do programu uvažovaného hudobného podujatia za posledných desať rokov. Veličinu X pozorujeme na $n = 39$ jednotkách výberového súboru (t.j. na $n = 39$ blokoch).

Friedmanov test spočíva v tom, že sa pozorovania na každom bloku usporiadajú zvlášť a určí sa poradie T_{ij} hodnoty x_{ij} v rámci i -teho bloku. Platí vzťah $\sum_{i=1}^n \sum_{j=1}^k T_{ij} = \frac{nk(k+1)}{2}$, ktorý môže slúžiť ako kontrola výpočtu. Testovacím kritériom je štatistika

$$Q = \frac{12}{nk(k+1)} \cdot \sum_{j=1}^k \left(\sum_{i=1}^n T_{ij} \right)^2 - 3n(k+1),$$

ktorá má za platnosti testovanej hypotézy H_0 asymptoticky χ^2 -rozdelenie s $k - 1$ stupňami voľnosti. Testovanú hypotézu H_0 zamietame na hladine významnosti α , ak hodnota

testovacieho kritéria Q prekročí kritickú hodnotu $\chi^2_\alpha(k-1)$. $\chi^2_\alpha(k-1)$ sú kritické hodnoty χ^2 - rozdelenia s $k-1$ stupňami voľnosti. Sú tabelované, čitateľ ich môže nájsť napríklad v tabuľke č. 12.7 publikácie [4].

Keďže štatistika Q má asymptoticky χ^2 - rozdelenie, predchádzajúci vzťah môžeme použiť len v prípade, že výbery majú veľký rozsah ($n > 20$). Pre malé hodnoty n a k ($n \leq 20$, $k \leq 12$) hodnotu testovacieho kritéria Q porovnáme s kritickou hodnotou Q_α Friedmanovho testu (tabuľka č. 12.18.a) v publikácii [4]). Testovanú hypotézu H_0 zamietame na hladine významnosti α , ak $Q \geq Q_\alpha$.

Ak zamietneme testovanú hypotézu H_0 v prospech alternatívnej hypotézy H_1 , ktorá znamená, že ošetrenia neprinášajú rovnaký efekt, ostáva nezodpovedaná otázka, ktoré súbory sa od seba štatisticky významne líšia. Na porovnanie rozdielov medzi jednotlivými súbormi použijeme Neményiho metódu mnohonásobného porovnávania pri Friedmanovom teste. Testovacím kritériom nulovej hypotézy H_0 , že j - te a m - te ošetrenie má rovnaký efekt, je

absolútna hodnota rozdielu súčtov poradí. Postupujeme nasledovne. Označíme $\sum_{j=1}^n T_{ij} = T_{.j}$ pre $j=1,2,\dots,k$ a vypočítame $|T_{.j} - T_{.m}|$ pre $j < m$. Ak je hodnota $|T_{.j} - T_{.m}|$ väčšia ako kritická tabuľková hodnota pre mnohonásobné porovnanie (tabuľka č. 12.18.b) v publikácii [3]), testovanú hypotézu, že j - te a m - te ošetrenie má rovnaký efekt, zamietneme na zvolenej hladine významnosti α . Asymptoticky sú kritické hodnoty pre tieto mnohonásobné porovnávanie dané vzťahom

$$q_{k,\infty}(\alpha) \cdot \sqrt{\frac{1}{12} \cdot nk(k+1)},$$

kde $q_{k,\infty}(\alpha)$ je kritická hodnota, ktorú nájdeme pre $\alpha = 0,05$ a $\alpha = 0,01$ v tabuľke č. 12.19 v publikácii [4]. Aproximáciu je možné použiť už pri $k > 5$.

Keďže realizácia štatistických výpočtov je časovo a numericky náročná, boli vyvinuté špeciálne štatistické softvéry, ktoré ich umožňujú ekonomizovať. V súčasnosti existuje veľké množstvo špeciálnych štatistických softvérov, čo je hlavne pre aplikovanú štatistiku veľkým prínosom. Štatistické programy umožňujú vyhodnotiť test aj bez znalosti kritických hodnôt testovacieho kritéria. Výstupná zostava počítača obsahuje okrem hodnoty testovacieho kritéria aj hodnotu pravdepodobnosti p . Je to pravdepodobnosť chyby, ktorej sa dopustíme, keď zamietneme testovanú hypotézu. Ak je vypočítaná hodnota pravdepodobnosti p dostatočne malá ($p < 0,05$ resp. $p < 0,01$), testovanú hypotézu H_0 zamietame (na hladine významnosti $\alpha = 0,05$ resp. $\alpha = 0,01$) v prospech alternatívnej hypotézy. V opačnom prípade hypotézu H_0 nemôžeme zamietnuť, pozorované rozdiely nie sú štatisticky významné.

3. Výsledky a diskusia

Keďže nemôžeme predpokladať normálne rozdelenie pozorovaného znaku, nulovú hypotézu H_0 : „Hudobné podujatia na Slovensku sa štatisticky významne nelíšia vzhľadom na zastúpenie diel slovenských skladateľov za posledných 10 rokov.“ budeme testovať Friedmanovým testom. Nulovú hypotézu budeme testovať oproti alternatívnej hypotéze H_1 : „Hudobné podujatia na Slovensku sa štatisticky významne líšia vzhľadom na zastúpenie diel slovenských skladateľov za posledných 10 rokov.“ Friedmanov test sme realizovali na počítači v programe STATISTICA. Po zadaní vstupných údajov sme dostali výstupnú zostavu

počítača (tabuľka č.1), v ktorej sú uvedené priemerné poradia, súčty poradí, priemery a smerodajné odchýlky. Výstupná zostava počítača obsahuje tiež hodnotu testovacieho kritéria χ^2 a hodnotu pravdepodobnosti p . Dostali sme nasledujúce výsledky: $\chi^2 = 109,0632$ a pravdepodobnosť $p < 0,00000$. Keďže hodnota p je dostatočne malá, testovanú hypotézu H_0 zamietame na hladine významnosti $\alpha = 0,01$ v prospech alternatívnej hypotézy. To znamená, že pozorované rozdiely sú štatisticky významné. Friedmanovým testom bolo potvrdené, že priemerné hodnoty pozorovaného znaku sú v jednotlivých hudobných podujatiach štatisticky významne rozdielne.

Tabuľka 1: Výstupná zostava počítača

	Priemerné poradie	Súčet poradí	Priemer	Smerodajná odchýlka
MSKE	4,026	157,000	0,410	0,850
FIB	6,846	267,000	1,897	2,382
CSST	7,718	301,000	2,154	1,631
CSMK	7,321	285,500	2,256	2,161
DHF JC	4,321	168,500	0,410	0,498
CSNL	4,244	165,500	0,462	0,720
KSDK	5,769	225,000	1,231	1,459
SP SHT	6,141	239,500	1,718	2,513
FSU KE	4,051	158,000	0,256	0,498
FNSH	4,564	178,000	0,564	0,940

V ďalšom nás bude zaujímať, ktoré hudobné podujatia sa navzájom líšia v zaradovaní klavírných diel slovenských hudobných skladateľov. To je predmetom ďalšej analýzy, v ktorej použijeme Neményiho metódu mnohonásobného porovnávania pri Friedmanovom teste. Vypočítame absolútne hodnoty rozdielov súčtov poradí pre každú dvojicu hudobných podujatí. Tieto hodnoty budeme porovnávať s kritickou hodnotou

$$q_{10, \infty}(0,05) \cdot \sqrt{\frac{1}{12} \cdot 39 \cdot 10 \cdot 11} = 4,47 \cdot 18,91 = 84,52.$$

Hodnotu $q_{10, \infty}(0,05) = 4,47$ sme vyhľadali v tabuľke č. 12.19 v publikácii [4]. Vypočítané absolútne hodnoty rozdielov súčtov poradí sú uvedené v nasledujúcej tabuľke. Testovanú hypotézu H_0 , že j -te a m -te hudobné podujatie sa navzájom nelíšia v zaradovaní klavírných diel slovenských hudobných skladateľov, zamietame na hladine významnosti $\alpha = 0,05$ v 15 prípadoch, označili sme ich *.

Tabuľka 2: Tabuľka absolútnych hodnôt rozdielov súčtov poradí

	MSKE	FIB	CSST	CSMK	DHF			SP	FSU
					JC	CSNL	KSDK	SHT	KE
MSKE									
FIB	110*								
CSST	144*	34							
CSMK	128,5*	18,5	15,5						
DHF JC	11,5	98,5*	132,5*	117*					
CSNL	8,5	101,5*	135,5*	120*	3				
KSDK	68	42	76	60,5	56,5	59,5			
SP SHT	82,5	27,5	61,5	46	71	74	14,5		
FSU KE	1	109*	143*	127,5*	10,5	7,5	67	81,5	
FNSH	21	89*	123*	107,5*	9,5	12,5	47	61,5	20

4. Záver

V predloženom článku je prezentovaný Friedmanov test a jeho praktické použitie pri vyhodnocovaní výsledkov výskumu z oblasti hudby. Výpočty sme realizovali v programe STATISTICA.

5. Literatúra

- [1]ANDĚL, J. 2003. Statistické metody. Matfyzpress, vydavateľstvo MFF KU Praha, 2003, ISBN 80-85863-27-8.
- [2]BOLDIŠOVÁ, V. 2010. Klavírna tvorba slovenských skladateľov a jej zastúpenie v hudobnom vzdelávaní na Slovensku (dizertačná práca). UKF v Nitre, 2010.
- [3]BOLDIŠOVÁ, V. 2010. Slovenská klavírna literatúra 20. storočia a jej prezentácia na vybraných hudobných podujatiach organizovaných na Slovensku. In: MUSICA ET EDUCATIO II. (Zborník príspevkov z konferencie doktorandov v Ružomberku, 2010), Ružomberok: Verbum, 2011, s. 140 – 148.
- [4]MARKECHOVÁ, D.– TIRPÁKOVÁ, A. 2011. Štatistické metódy a ich aplikácie. FPV UKF v Nitre, 2011, ISBN 978-80-8094-807-8.

Adresa autorov:

Dagmar Markechová, Doc., RNDr., CSc.
 Trieda A. Hlinku 1
 949 74 Nitra
 dmarkechova@ukf.sk

Využitie softvéru Mathematica pri názornosti kriviek¹ Using of Mathematica software by visualization of curve

Ladislav Mura

Abstract:

In this paper we presented how to visualize a curve with help of Mathematica software. Two techniques are presented. A symbolic way and an engineering way of visualizing a curve are shown. Number of examples of Mathematica-codes are followed by its graphical outputs both in two-dimensional and three-dimensional space. Static curves are created by functions ParametricPlot, ParametricPlot3D and Graphics3D. For creating a dynamic curve function Manipulate is added.

Key words:

Curve, visualization, Mathematica software

Kľúčové slová:

Krivka, názornosť, program Mathematica

JEL classification:

C88, C60

1. Introduction

Mathematica is a computational software program used in scientific, engineering, and mathematical fields and other areas of technical computing. Mathematica can connect to many outside programs. You can use Mathematica's rich programming language to read and write to other supported programming languages. Mathematica is a truly fantastic tool which allows you to analyze and visualize large amounts of data, and many other things which can be very helpful. [2]

Mathematica software is a large scale product for engineering, science, economics and math education. Mathematica provides powerful functions that automate the process of creating representations of structured and unstructured data. In this article is shown a use of Mathematica for the purpose of visualization of a curve.[3]

2. Visualization in 2D

2.1 Static curve in 2D

Firstly we will show how to create a static curve with help of built-in Mathematica function *ParametricPlot*. The syntax of *ParametricPlot* is following [1]:

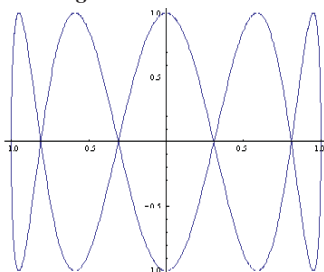
```
ParametricPlot[{fx, fy}, {u, Umin, Umax}]
```

generates a parametric plot of a curve with x and y coordinates f_x and f_y as a function of u .

Mathematica-code:

```
Clear[c]
c[t_]:= {Cos[t], Sin[5t]}
t1=0; t2=2 Pi;
ParametricPlot[c[t], {t, t1, t2}]
```

¹ This paper was supported by internal Grant „Internationalization of small and medium enterprises in chosen region“ of Dubnica Institute of Technology in Dubnica nad Váhom

Figure 1 Static curve

Source: output from software Mathematica

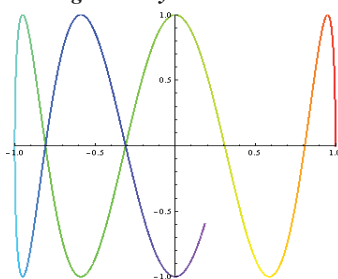
The plot range was in this case selected automatically. There are 21 basic options that enable us to set up the graphical output. In next examples we will use options: *PlotRange*, *PlotStyle* and *ColorFunction*.

2.2 Dynamic curve in 2D

Now we will create a dynamic curve with help of built-in function *Manipulate*. The two following Mathematica codes differ only in options and therefore we get two different outputs. First code corresponds with Fig.2 - left image, second code with Figure 2 - right image.

Mathematica-code:

```
Manipulate[ParametricPlot[c[t],{t, t1, T},
PlotRange->{{-1, 1},{-1, 1}}, PlotStyle->Thick,
ColorFunction->Function[{x, y, t}, Hue[t/(2 Pi)]],
ColorFunctionScaling->False],
{T, t1+0.1, t2}]
Manipulate[ ParametricPlot[c[t],{t, t1, T},
PlotRange->{{-1, 1}, {-1, 1}},
PlotStyle->Thick],
{T, t1+0.1, t2}]
```

Figure 2 Dynamic curve

Source: output from software Mathematica

2.3 Dynamic Curve in 2D Together with a moving point and an arrow

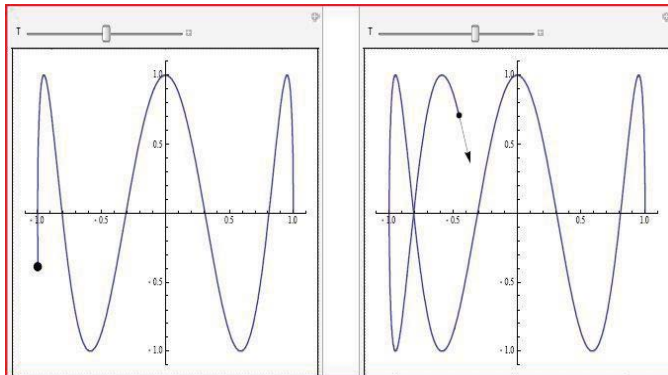
The parameter *T* in function *Manipulate* gives us a possibility to "draw" the curve and see how it develops in time. To make the image more descriptive, we can add a moving point

and an arrow to the ending of our curve. The arrow follows direction of a tangent line. The `PlotRange` is set to `{-1.1, 1.1}` so that the arrow is visible during its whole trace.

Mathematica-code:

```
Manipulate[
Show[
ParametricPlot[c[t],{t, t1, T},
PlotRange->{{-1.1, 1.1}, {-1.1, 1.1}},
PlotStyle->Thick],
Graphics[{Black,PointSize[0.02], Point[c[T]]}, Axes->True]
{T, t1+0.01, t2}]
Manipulate[
Show[
ParametricPlot[c[t],{t, t1, T},
PlotRange->{{-1.1, 1.1}, {-1.1, 1.1}},
PlotStyle->Thick],
Graphics[{Black,PointSize[0.02], Point[c[T]],
Arrow[{c[T],c[T]+0.1c'[T]}]}, Axes->True]
{T, t1+0.01, t2}]
```

Figure 3 Dynamic curve together with a moving point (on the left) and dynamic curve together with a moving point and an arrow (on the right)



Source: output from software Mathematica

3. Visualization in 3D

3.1 Static curve in 3D

To create a static curve in 3D we will use a built-in function *ParametricPlot3D* with syntax [4]:

```
ParametricPlot3D[{fx, fy, fz}, {u, Umin, Umax}]
```

produces a three-dimensional space curve parametrized by a variable u which runs from $Umin$ to $Umax$.

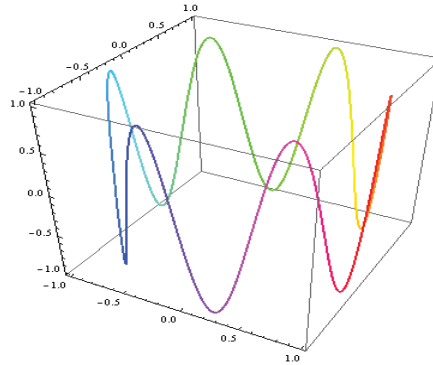
Mathematica-code:

```
Clear[cu]
cu[t_]:= {Cos[t], Sin[t], Sin[6t]}
t1=0; t2=2Pi;
ParametricPlot3D[cu[t],{t, t1, t2}, PlotStyle->Thick]
```



```
ParametricPlot3D[cu[t],{t,t1,t2},PlotStyle->Thick,
ColorFunction->Function[{x,y,z,t},Hue[t]]]
```

Figure 3 Static curve in 3D, function prescription is known



Source: output from software Mathematica

3.2 Creating a Data Set

In all examples until now, we created a curve with knowledge of a function ($c[t]$ in 2D and $cu[t]$ in 3D). It is so called "symbolic way" of creating a curve. However, in many cases no functionality is known and we want to create a curve on a base of a data set. In this case we will use a Mathematica built-in function *Graphics3D* and we will call it "engineering way". Usually we get the data set as a result of an experiment. Here we will create a data set with help of previously used function $cu[t]$.

Mathematica-code:

```
data=N[Table[Flatten[{t,cu[t]}],{t,t1,t2,(t2-t1)/400}]];
(*Flatten changes order to: t,x,y,z, *)
Data set:
{{0.,1.,0.,0.},{0.015708,0.999877,0.0157073,0.0941083},{0.0314159,0.999507,0.0314108,0.187381},
{0.0471239,0.99889,0.0471065,0.278991},{0.0628319,0.998027,0.0627905,0.368125},
{0.0785398,0.996917,0.0784591,0.457269},
{0.0941065,-0.278991},{6.25177,0.999507,-0.0314108,-0.187381},
{6.26748,0.999877,-0.0157073,-0.0941083},
{6.28319,1.,0.,0.}}
```

To create a static curve in 3D based on a data set we will use a built-in function *Graphics3D* with syntax [5]:

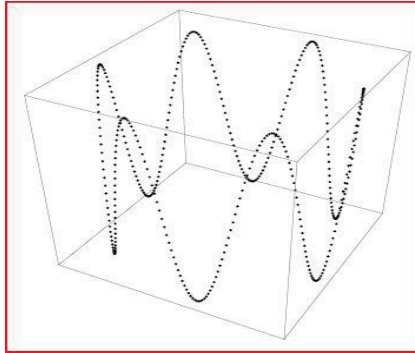
Graphics3D[*primitives, options*]

represents a three-dimensional graphical image

Mathematica-code:

```
Graphics3D[
Point[data/.{s_,u_,v_,w_}->{u,v,w}]
]
Graphics3D[
data/.{s_,u_,v_,w_}->{Hue[s],Point[{u,v,w}]]]
```

Figure 4 *Static curve in 3D, based on a data set*



Source: output from software Mathematica

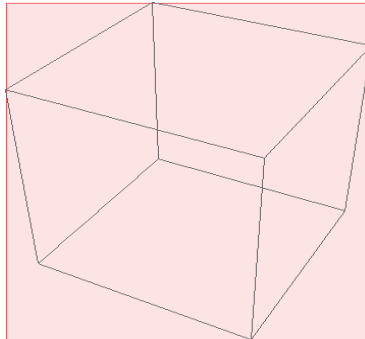
Each point on the graph corresponds with one item of our data set. Drawing the curve one point after another enables function Manipulate which we show in next paragraph.

3.3 Dynamic curve in 3D

Mathematica-code:

```
Manipulate[
Graphics3D[
Take[data,k]/.{s_, u_, v_, w_}->
{Hue[s/(2 Pi)], Point[{u, v, w}]},
PlotRange->{{-1, 1}, {-1, 1}, {-1, 1}}
],
{k, 1, Length[data], 1}]
```

Figure 5 *Dynamic curve in 3D, fails for k=4*



Source: output from software Mathematica

As we see on the figure 5, this code fails when we select $k = 4$. Here "k" is a sequential number of a displayed point . We repair this defect by declaring a variable s a real number (s_Real).

Mathematica-code:

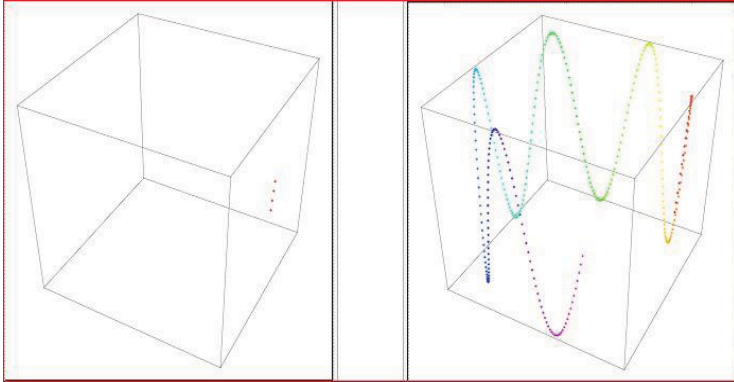
```
Manipulate[
Graphics3D[
```

```

Take[data, k]/. {s_Real, u_, v_, w_}->
{Hue[s/(2 Pi)], Point[{u, v, w}]},
PlotRange -> {{-1, 1}, {-1, 1}, {-1, 1}}
],
{k, 1, Length[data], 1}]

```

Figure 6 *Dynamic curve in 3D, works for all k*



Source: output from software Mathematica

4. Conclusion

Mathematica software is a useful instrument for scientists as well as math teachers. Even if used only in a passive way, it offers thousands of ready-made examples, that can be downloaded for free together with explanation text and source-code from *wolfram.com* pages. Creating own demonstrations with help of built-in functions is very easy. Lessons can become more attractive and understandable for students with help of Mathematica.

5. References

- [1] FRI ŽU. Program Mathematica. [online] [cit. 2011-03-09] Available from: <http://www.kis.fri.uniza.sk/index.php/en/departament/we-ecommend/mathematica/252>
- [2] MURA, L. Zhlukovanie ako štatistická analýza v ekonomických vedách. In: Forum Statisticum Slovacum, No. 5/2010, pp. 165 – 170. ISSN 1336-7420 [online] [cit. 2011-03-09] Available from: <http://www.ssds.sk/casopis/archiv/2010/fss0510.pdf>
- [3] Software Mathematica, Parametric Plot. [online] [cit. 2011-04-10] Available from: <http://reference.wolfram.com/mathematica/ref/ParametricPlot.html>
- [4] Software Mathematica, Parametric Plot3D. [online] [cit. 2011-04-10] Available from: <http://reference.wolfram.com/mathematica/ref/ParametricPlot3D.html>
- [5] Software Mathematica, Graphics3D.[online] [cit. 2011-04-10] Available from: <http://reference.wolfram.com/mathematica/ref/Graphics3D.html>

Contact:

Ladislav Mura, Ing. et Bc., PhD.
 Ústav odborných predmetov a IT
 Dubnický technologický inštitút v Dubnici nad Váhom
 Ul. Sládkovičova 533/20, 018 41 Dubnica nad Váhom
 E mail: ladislav.mura@gmail.com

Jednorozměrná kalibrace s kvadratickou kalibrační funkcí One-dimensional calibration with quadratic calibration function

Kateřina Myšková

Abstract: The article deals with one-dimensional calibration assuming a quadratic calibration function. Corresponding model as well as estimates of the unknown parameters is derived. Knowledge of measurement models is used here, namely the incomplete direct measurement model with second type conditions for first order parameters. A small simulation study on the influence of the calibration function parameters is performed.

Key words: one-dimensional calibration, quadratic calibration function, estimate of parameters, simulation study.

Klíčové slová: jednorozměrná kalibrace, kvadratická kalibrační funkce, odhady parametrů, simulační studie.

JEL classification: C39

1. Úvod

Jsou situace, kdy se předpoklad o lineárním vztahu mezi měřeními zdá nedostačující. Jedním z možných zobecnění je předpokládat kalibrační funkci v kvadratickém tvaru. Je to smysluplný předpoklad, protože se vzrůstajícími naměřenými hodnotami by nepřesnost mohla růst více (nebo méně) než lineárně. Např. u měření vlhkosti semen lze předpokládat kvadratickou kalibrační funkci.

Proto se v tomto článku budeme zabývat sestavením modelu jednorozměrné kalibrace (obě měření jsou zatížena chybami) za předpokladu, že vztah mezi měřeními je v kvadratickém tvaru, a odvozením odhadů neznámých parametrů modelu.

2. Předpoklady modelu

Nechť $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ je vektor náhodných veličin „prvních“ měření, pro které platí

$$X_i \sim N(\mu_i, \sigma_x^2) \text{ pro } i = 1, \dots, n.$$

Nechť $\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$ je vektor náhodných veličin „druhých“ měření, pro které platí

$$Y_i \sim N(v_i, \sigma_y^2) \text{ pro } i = 1, \dots, n.$$

Označme $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ a předpokládejme nekorelovanost prvních a druhých měření

i měření mezi objekty. Potom „celkový“ model měření zapíšeme ve tvaru

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N \left[\begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{v} \end{pmatrix}, \begin{pmatrix} \sigma_x^2 \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{0}_n & \sigma_y^2 \mathbf{I}_n \end{pmatrix} \right].$$

Dále předpokládáme, že vztah mezi skutečnými (bezchybnými) hodnotami měření (kalibrační funkce) je v kvadratickém tvaru:

$$v_i = a + b\mu_i + c\mu_i^2 \text{ pro } i = 1, \dots, n$$

neboli

$$\mathbf{v} = a\mathbf{1}_n + b\boldsymbol{\mu} + c\boldsymbol{\mu}^2,$$

kde $\boldsymbol{\mu}^2 = (\mu_1^2, \dots, \mu_n^2)'$, $\mathbf{1}_n = (1, \dots, 1)' \in R^n$.

Podmínky linearizujeme pomocí Taylorova rozvoje v okolí $b_0, c_0, \boldsymbol{\mu}_0 = (\mu_{10}, \dots, \mu_{n0})'$. Zanedbáním členů druhého a vyšších řádů dostáváme

$$v_i \doteq a + b_0\mu_{i0} + c_0\mu_{i0}^2 + \mu_{i0}(b - b_0) + \mu_{i0}^2(c - c_0) + (b_0 + 2c_0\mu_{i0})(\mu_i - \mu_{i0}).$$

Po úpravě

$$v_i \doteq a + \mu_{i0}b + \mu_{i0}^2c + b_0\delta\mu_i + 2c_0\mu_{i0}\delta\mu_i$$

neboli

$$\mathbf{v} \doteq \mathbf{1}_n a + \boldsymbol{\mu}_0 b + \boldsymbol{\mu}_0^2 c + \text{diag}(b_0 \mathbf{1}_n + 2c_0 \boldsymbol{\mu}_0) \delta \boldsymbol{\mu},$$

kde $\delta\mu_i = \mu_i - \mu_{i0}$ pro $i = 1, \dots, n$, $\delta\boldsymbol{\mu} = (\delta\mu_1, \dots, \delta\mu_n)'$ a $\text{diag}(b_0 \mathbf{1}_n + 2c_0 \boldsymbol{\mu}_0)$ je diagonální matice, jejíž prvky na diagonále odpovídají složkám vektoru $b_0 \mathbf{1}_n + 2c_0 \boldsymbol{\mu}_0$.

V závislosti na vektorech neznámých parametrů můžeme podmínky přepsat do tvaru

$$\left(\text{diag}(b_0 \mathbf{1}_n + 2c_0 \boldsymbol{\mu}_0), -\mathbf{I}_n \right) \begin{pmatrix} \delta \boldsymbol{\mu} \\ \mathbf{v} \end{pmatrix} + (\mathbf{1}_n, \boldsymbol{\mu}_0, \boldsymbol{\mu}_0^2) \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{0}.$$

Celkově tedy dostáváme model

$$\begin{pmatrix} \mathbf{X} - \boldsymbol{\mu}_0 \\ \mathbf{Y} \end{pmatrix} \sim N \left[\begin{pmatrix} \delta \boldsymbol{\mu} \\ \mathbf{v} \end{pmatrix}, \begin{pmatrix} \sigma_x^2 \mathbf{I}_n & \mathbf{0}_n \\ \mathbf{0}_n & \sigma_y^2 \mathbf{I}_n \end{pmatrix} \right]$$

s podmínkami na parametry

$$\left(\text{diag}(b_0 \mathbf{1}_n + 2c_0 \boldsymbol{\mu}_0), -\mathbf{I}_n \right) \begin{pmatrix} \delta \boldsymbol{\mu} \\ \mathbf{v} \end{pmatrix} + (\mathbf{1}_n, \boldsymbol{\mu}_0, \boldsymbol{\mu}_0^2) \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbf{0}.$$

3. Odhady parametrů

Výše uvedený model je speciálním případem modelu neúplného přímého měření s podmínkami II. typu na parametry 1. řádu (viz např. [1, str. 129]). Na základě věty uvedené tamtéž vyjádříme odhady neznámých parametrů. Označme

$\mathbf{W} = \text{diag}(\sigma_x^2 (b_0 \mathbf{1}_n + 2c_0 \boldsymbol{\mu}_0)^2 + \sigma_y^2 \mathbf{1}_n)$, $\mathbf{A} = (\mathbf{1}_n, \boldsymbol{\mu}_0, \boldsymbol{\mu}_0^2)$, $\mathbf{D} = \text{diag}(b_0 \mathbf{1}_n + 2c_0 \boldsymbol{\mu}_0)$, potom

$$\hat{\boldsymbol{\mu}} = \mathbf{X} - \sigma_x^2 \mathbf{D}(\mathbf{P}_A \mathbf{W} \mathbf{P}_A)^+ (\mathbf{D}(\mathbf{X} - \boldsymbol{\mu}_0) - \mathbf{Y}),$$

$$\hat{\mathbf{v}} = \mathbf{Y} + \sigma_y^2 (\mathbf{P}_A \mathbf{W} \mathbf{P}_A)^+ (\mathbf{D}(\mathbf{X} - \boldsymbol{\mu}_0) - \mathbf{Y}),$$

$$\begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix} = -(\mathbf{A}' \mathbf{W}^{-1} \mathbf{A})^{-1} \mathbf{A}' \mathbf{W}^{-1} (\mathbf{D}(\mathbf{X} - \boldsymbol{\mu}_0) - \mathbf{Y}),$$

kte $(\mathbf{P}_A \mathbf{W} \mathbf{P}_A)^+ = \mathbf{W}^{-1} - \mathbf{W}^{-1} \mathbf{A}(\mathbf{A}' \mathbf{W}^{-1} \mathbf{A})^{-1} \mathbf{A}' \mathbf{W}^{-1}$. Jedná se o Mooreovu-Penroseovu pseudoinverzi (viz např. [2, str. 47]). Dále kovarianční matice odhadů jsou

$$\text{var} \begin{pmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} \sigma_x^2 \mathbf{I}_n - \sigma_x^4 \mathbf{D}(\mathbf{P}_A \mathbf{W} \mathbf{P}_A)^+ \mathbf{D} & \sigma_x^2 \sigma_y^2 \mathbf{D}(\mathbf{P}_A \mathbf{W} \mathbf{P}_A)^+ \\ \sigma_x^2 \sigma_y^2 (\mathbf{P}_A \mathbf{W} \mathbf{P}_A)^+ \mathbf{D} & \sigma_y^2 \mathbf{I}_n - \sigma_y^4 (\mathbf{P}_A \mathbf{W} \mathbf{P}_A)^+ \end{pmatrix},$$

$$\text{var} \begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix} = -(\mathbf{A}' \mathbf{W}^{-1} \mathbf{A})^{-1},$$

$$\text{cov} \left(\begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix}, \begin{pmatrix} \hat{\boldsymbol{\mu}} \\ \hat{\mathbf{v}} \end{pmatrix} \right) = (\mathbf{A}' \mathbf{W}^{-1} \mathbf{A})^{-1} \mathbf{A}' \mathbf{W}^{-1} (\sigma_x^2 \mathbf{D}, -\sigma_y^2 \mathbf{I}_n).$$

4. Simulační studie

Simulační studie byla zaměřena na zkoumání kvality odhadu parametrů kalibrační funkce v závislosti na jejich hodnotách. K výpočtům jsme využili výpočetní systém Matlab R2008. Možné vychýlení jsme měřili střední chybou (ME) a kvalitu odhadu střední kvadratickou chybou (MSE). Pro parametr a uvádíme vzorec, pro ostatní parametry se použijí analogická vyjádření: $ME = \frac{1}{n} \sum_{i=1}^n (\hat{a}_i - a)$, $MSE = \frac{1}{n} \sum_{i=1}^n (\hat{a}_i - a)^2$. Parametry kalibrační funkce byly voleny 10 způsobů (viz Tabulka 1). Realizace „prvních“ měření byly generovány za předpokladu normálního rozdělení s vektorem středních hodnot $\boldsymbol{\mu} = (0, 1, \dots, 25)'$ a kovariancí $\sigma_x^2 = 0,1$, realizace „druhých“ měření byly generovány za předpokladu normálního rozdělení s vektorem středních hodnot $\mathbf{v} = a \mathbf{1}_n + b \boldsymbol{\mu} + c \boldsymbol{\mu}^2$ a kovariancí $\sigma_y^2 = 0,5$. Pro každou kombinaci jsme generovali 10 000 realizací.

Za počáteční odhad středních hodnot měření jsme zvolili realizace měření, tedy $\boldsymbol{\mu}_0 = \mathbf{X}$. Počáteční odhady parametrů kalibrační funkce jsme odvodili z modelu

$$\mathbf{Y} \sim \left[(\mathbf{1}_n, \mathbf{X}, \mathbf{X}^2) \begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix}, \mathbf{I}_n \right].$$

V následující tabulce jsou uvedeny výsledky.

Tabulka 1: Výsledky malé simulační studie

	a	ME(a)	MSE(a)	b	ME(b)	MSE(b)	c	ME(c)	MSE(c)
1.	0	-0,025275	0,156436	0	0,006817	0,024990	1	-0,000300	0,000072
2.	0	-0,018708	0,118749	0	0,003307	0,011118	0,5	-0,000113	0,000027
3.	0	-0,056438	0,248261	0	0,019323	0,212141	5	-0,000753	0,000862
4.	-5	-0,028323	0,154665	0	0,006804	0,024220	1	-0,000237	0,000069
5.	5	-0,025439	0,158895	0	0,006619	0,025665	1	-0,000237	0,000073
6.	25	-0,004252	0,323746	-10	-0,004756	0,019648	1	0,000323	0,000046
7.	100	0,031415	0,827185	-20	-0,009731	0,025418	1	0,000450	0,000044
8.	-125	-0,052679	1,619678	30	0,011842	0,033751	-1	-0,000449	0,000043
9.	-375	0,005312	0,315799	10	0,004268	0,019561	-1	-0,000303	0,000047
10	1	-0,202719	0,449246	2	0,096242	2,951186	20	-0,004606	0,013078

5. Závěr

Z výsledků simulace vyplývá, že nejlepších výsledků (ve smyslu ME i MSE) je dosaženo pro parametr c . Při konstantních hodnotách parametrů a , b je vidět, že se vzrůstající hodnotou parametru c roste jeho vychýlení a snižuje se kvalita (1., 2., 3. řádek). Změna hodnoty konstantního členu a při pevných hodnotách parametrů b , c nemá vliv ani na vychýlení ani na kvalitu odhadu (1., 4., 5. řádek). S rostoucí absolutní hodnotou parametru b , resp. s posunem vrcholu paraboly do „měřeného rozsahu dat“ roste vychýlení a snižuje se kvalita (5., 6., 7., 8. řádek). Z uvedených výsledků také vyplývá, že rostoucí hodnota parametru c ovlivňuje vychýlení (zvyšuje) i kvalitu (snižuje) odhadu ostatních parametrů.

Při psaní a odladování programové implementace byl zaznamenán značný vliv kovariačních matic (přesností měření) a to, že se vzrůstajícím rozptylem rostou i nepřesnosti odhadů.

6. Literatura

- [1] KUBÁČEK, L. – KUBÁČKOVÁ, L. 2000. Statistika a metrologie. Olomouc: Univerzita Palackého v Olomouci, 2000. 307 s. ISBN 80-244-0093-6.
- [2] RAO, C. R. 1978. Lineární metody statistické indukce a jejich aplikace. Praha: Academia, 1978. 668 s. ISBN 104-21-852.

Adresa autora:

Kateřina Myšková, Mgr. Ph.D.
 ÚSOV PEF MENDELU
 Zemědělská 1
 613 00 Brno
 myskova@mendelu.cz

Zmeny vo vývoji evidovanej miery nezamestnanosti v Nitrianskom kraji Changes in the development registered unemployment rate in the Nitra Region

Peter Obtulovič, Zuzana Poláková

Abstract: The importance of modern forecasting approaches Exponential Smoothing and Box - Jenkins methodology is constantly increasing. The focus of their use, we see mainly Sup-
pose replying to the future development of such economic indicators, development of which is marked by great volatility and seasonality.

Key words: unemployment, Box - Jenkins methodology, forecast, time series, forecasting model

Kľúčové slová: nezamestnanosť, Box – Jenkinsova metodológia, predpoveď, časové rady, modely

JEL classification: C13, C22,

1. Úvod

Nezamestnanosť je prirodzeným fenoménom a atribútom slobodnej spoločnosti založenej na trhovom hospodárstve a demokracii. Jej nekontrolovateľný vývoj spôsobujúci masový charakter však vyvoláva nielen vážne ekonomické, ale aj sociálne problémy (rozpad rodiny, narušené mentálne i fyzické zdravie a rôzne sociálno-patologické javy). Štatistiky potvrdzujú, že najväčší podiel nezamestnaných tvoria občania s neúplným či žiadnym vzdelaním, preto predpokladom ich presadenia sa na trhu práce je hlavne ich vlastná iniciatíva vedúca k rozšíreniu vzdelania či schopností, ktoré sú momentálne na trhu najžiadanejšie. Aktuálnosť tejto problematiky sa zvyšuje aj v súvislosti s globálnou ekonomickou krízou, ktorá od jesene roku 2008 výrazne ovplyvnila i otvorenú ekonomiku Slovenska.

V súvislosti s globálnou ekonomickou krízou dochádza na Slovensku od posledného štvrtroka 2008 k zmene vo vývoji evidovanej miery nezamestnanosti. Nárast miery nezamestnanosti v tomto krízovom období je u všetkých skupín obyvateľov ako aj vo všetkých krajoch Slovenska. Z hľadiska odvetví národného hospodárstva globálna ekonomická kríza zasiahla najmä sekundárny sektor, čím získal tento sektor prvenstvo v úrovni nezamestnanosti pred primárnym sektorom. S pretrvávajúcimi rozdielmi v zložení obyvateľstva podľa veku, vzdelania ako aj s rozdielmi v ekonomickej úrovni regiónov Slovenska súvisia aj rozdiely v úrovni nezamestnanosti.

Využívanie matematicko štatistických postupov pri rôznych formách optimalizačných postupov má už v súčasnom období nezastupiteľné miesto. Tak je tomu aj pri konštrukcii predpovedí resp. prognóz takých ukazovateľov, ktoré sa vyvíjajú pod vplyvom rôznych faktorov súčasného turbulentného ekonomického prostredia u nás.

Počtom obyvateľov sa Nitriansky kraj zaraďuje na 3. miesto v medzi krajskom porovnaní s podielom 13 % na úhrne SR. Z celkového počtu obyvateľov je 51,6 % žien. S hustotou osídlenia 111,2 obyvateľov na km² je piatym najobývanejším krajom SR. V Nitrianskom kraji sa nachádza 354 obcí, z ktorých má 15 štatút mesta. V mestách žije 329 054 obyvateľov, čo je 46,6 % podiel zo všetkých obyvateľov. V Nitrianskom kraji je možné pozorovať výrazné zmeny v demografickom vývoji, ktoré sú odrazom ekonomickej a sociálnej situácie kraja. Prejavujú sa dlhodobejšie tendencie spomaľovania reprodukcie obyvateľstva a znižovania prirodzeného prírastku obyvateľstva. Z ostatných demografických procesov dochádza k zni-

žovaniu sobášnosti a rozvodovosti. Pokračuje proces starnutia obyvateľstva. Priemerný vek obyvateľov kraja v roku 2009 bol 39,83 rokov a index starnutia dosiahol hodnotu 98,89 (podľa metodiky EÚ).

2. Materiál a metódy

Využitelnosť adaptívnych prístupov k modelovaniu vývoja časových radov bola empiricky overená na súbore mesačných časových radov evidovanej miery nezamestnanosti v období január 1997 až február 2011 za jednotlivé okresy Nitrianskeho kraja ako aj za celý kraj. Takmer všetky časové rady mali veľmi zložitý a nepravidelný priebeh s množstvom výkyvov. Pre prognózu evidovanej miery nezamestnanosti sme zvolili časový horizont o dĺžke 12 mesiacov, t.j. marec 2011 až február 2012. Kvalitu prognóz sme posudzovali pomocou priemernej relatívnej chyby predpovede MAPE. Pre všetky analyzované časové rady sme prepočítali 42 najčastejšie používaných adaptívnych modelov (preddefinovaných programovým produktom SAS), z ktorých boli vybrané tri najlepšie modely pre výpočet kombinovanej prognózy, na základe rozhodovacieho kritéria, za ktoré sme zvolili MAPE. Prognózy boli použité na výpočet očakávaných hodnôt mier evidovanej nezamestnanosti časových radov za jednotlivé okresy Nitrianskeho kraja.

3. Výsledky a diskusia

V príspevku sme analyzovali celkom 8 časových radov evidovanej miery nezamestnanosti. Vzhľadom na rozsah príspevku v ďalšej časti detailne popíšeme len prognózu miery evidovanej nezamestnanosti Nitrianskeho kraja. Vývoj časového radu evidovanej miery nezamestnanosti v období Jan1997 až Feb2011, ako aj vypočítanú prognózu na obdobie Mar2011 až Feb2012 uvádzame súhrnne v grafe 1.

Vývoj priemernej evidovanej miery nezamestnanosti v Nitrianskom kraji mal od januára 1997 do januára 2002 rastúci charakter. Od roku 2003 začala nezamestnanosť v Nitrianskom kraji postupne klesať. Od novembra 2008 v dôsledku dopadu krízy nezamestnanosť rástla. Z prognózy na obdobie marec 2011 až február 2012 vyplýva, že evidovaná miera nezamestnanosti by mala opäť ku koncu prognózovaného obdobia rásť až k hodnote 14 %. Vypočítané výsledky prognóz evidovanej miery nezamestnanosti za Nitriansky kraj a všetky okresy Nitrianskeho kraja uvádzame súhrnne v tabuľke 1.

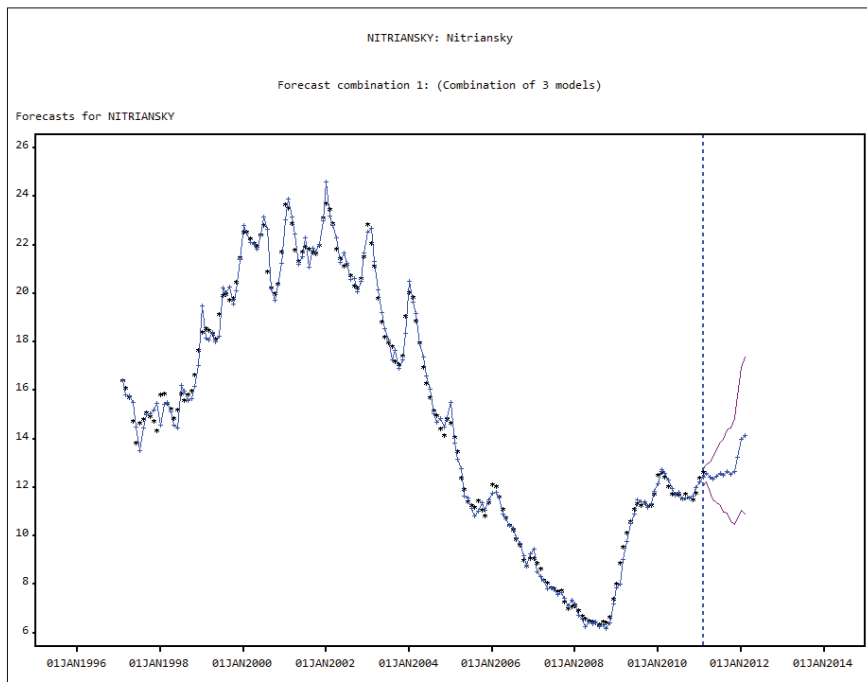
Výsledky empirickej analýzy poukázali na možnosti využitia adaptívnych modelov pri konštrukcii krátkodobých predpovedí časových radov ekonomických ukazovateľov. Exaktné poznatky o budúcom vývoji hodnôt časového radu sú pre ekonomické prostredie jednou z dôležitých informácií ovplyvňujúcich strategické rozhodovanie podnikateľských subjektov.

Ak zhrnieme poznatky, ktoré sme získali pri výstavbe adaptívnych modelov miery nezamestnanosti môžeme ich sformulovať do nasledovných záverečných vyjadrení:

- ♦ výhodou adaptívnych prístupov modelovania časových radov ekonomických ukazovateľov je, že sú flexibilné a rýchlo sa adaptujú na zmeny vo vývoji časového radu. Stochasticky modelujú nielen trendovú zložku, ale aj sezónnu zložku časových radov

- ♦ adaptívne modely sú schopné popísať aj také časové rady ekonomických ukazovateľov, ktoré sa vyvíjajú nepravidelne, s meniacim sa trendom, ale aj s meniacou sa sezónnou zložkou. Takéto časové rady sú práve typické pre naše nestabilné ekonomické prostredie.

- ♦ praktické aplikácie a z nich získané kvalitné predpovede potvrdzujú vhodnosť adaptívnych prístupov k modelovaniu časových radov, kedy v porovnaní s inými metódami dosahujú najkvalitnejšie výsledky, ale často strácame možnosť jednoduchej interpretácie vypočítaných parametrov výsledných modelov



Graf 1 Vývoj a prognóza priemernej evidovanej miery nezamestnanosti za Nitriansky kraj, v % (výstup programu SAS)

Zdroj: Vlastné výpočty

Tabuľka 1 Vývoj evidovanej miery nezamestnanosti a hodnoty MAPE v prognózovanom období Marec2011 - Feb2011 pre Nitriansky kraj a jeho okresy v %.

Kraj	Nitriansky kraj	okres Komárno	okres Levice	okres Nitra	okres Nové Zámky	okres Šaľa	okres Topoľčany	okres Zlaté Moravce
III-11	12,57	16,99	15,61	8,07	13,16	10,58	11,17	10,91
IV-11	12,43	16,72	15,20	7,77	12,83	10,43	10,89	10,59
V-11	12,36	16,42	14,89	7,76	12,63	10,26	10,66	10,44
VI-11	12,44	16,24	14,87	7,88	12,57	10,27	10,70	10,46
VII-11	12,56	16,20	15,05	8,03	12,63	10,44	10,89	10,52
VIII-11	12,49	16,00	14,96	7,98	12,46	10,19	10,81	10,44
IX-11	12,64	16,01	15,00	8,05	12,54	10,34	10,85	10,52
X-11	12,53	15,92	14,88	7,90	12,43	10,21	10,66	10,57
XI-11	12,65	16,17	15,17	7,92	12,62	10,15	10,69	10,67
XII-11	13,23	17,02	15,78	8,35	13,13	10,50	10,96	11,09
I-12	14,00	18,05	16,56	8,86	13,81	10,95	11,38	11,91
II-12	14,14	18,46	16,44	8,87	13,84	10,99	11,35	11,76
MAPE	1,95	2,41	2,42	2,61	2,28	2,55	2,42	3,43

Zdroj: Vlastné výpočty

Z hľadiska nezamestnanosti v jednotlivých okresoch v prognózovanom období nepredpokladáme žiadne prudké zvýšenie, či zníženie nezamestnanosti. Úroveň evidovanej miery nezamestnanosti v tomto regióne je relatívne stabilizovaná a hlavne koncom roka je možné očakávať mierny nárast evidovanej nezamestnanosti.

V okrese Komárno môžeme predpokladať zníženie priemernej nezamestnanosti v prognózovanom období z pôvodnej hodnoty 16,99 % v marci 2011 na 15,92 % v októbri 2011. Od novembra 2011 do februára 2012 očakávame nárast hodnôt priemernej nezamestnanosti až na konečnú úroveň 18,46 % vo februári 2012. V rámci okresu Levice predpokladáme, že počas prognózovaného obdobia dôjde k poklesu úrovne miery nezamestnanosti. Nezamestnanosť by mala do októbra 2011 klesnúť na úroveň 14,88 %, čím sa zníži oproti úrovni v marci 2011 len o 0,73 %. Zvýšenie počas prognózovaného obdobia očakávame v zimných mesiacoch až na hodnotu 16,44 %. Tretím okresom v poradí s najvyššou mierou evidovanej nezamestnanosti je okres Nové Zámky, v ktorom očakávame zníženie evidovanej miery nezamestnanosti v prognózovanom období z 13,16 % na 12,43 % v októbri 2011. V zimných mesiacoch podobne ako v predchádzajúcich dvoch okresoch, očakávame jej nárast až na úroveň 13,84 %. Z uvedeného je zrejmé, že naďalej najnižšia miera nezamestnanosti je v okresoch Nitra a Šaľa.

Štát sociálnymi dávkami úspešne konkuruje nízkym mzdám, a preto kategóriu dlhodobo nezamestnaných v okresoch s najvyššou mierou evidovanej nezamestnanosti tvoria väčšinou ľudia bez vzdelania, ktorí sa za vyššiu mzdu zamestnať objektívne nedokážu. Úrovni dávok poskytovaných v nezamestnanosti vo vzťahu k úrovni príjmov zo závislej činnosti je nevyhnutné venovať dostatočnú pozornosť aj v nasledujúcom období. Dnešný systém akúkoľvek motiváciu pracovať za nízku mzdu definitívne zabíja. Mali by sa preto vytvoriť podmienky pre regulovanie výšky dávok poskytovaných v nezamestnanosti a dávok sociálnej pomoci v závislosti na pracovných príjmoch tak, aby úroveň týchto dávok nútila zaujímať sa o zapojenie do pracovného procesu.

Mnohí dlhodobo nezamestnaní sú už takpovediac zmierení so svojou situáciou a v podstate už ani neprejavujú záujem o začlenenie sa do pracovného procesu. Nástup do zamestnania by pre nich predstavoval veľkú životnú zmenu. Preto by bolo potrebné prijať opatrenia, ktoré by nedovolili nezamestnaným zostať nečinnými, ale naopak neustále ich motivovať, rozvíjať ich schopnosti a dávať im pocit potrebnosti pre spoločnosť. Tiež odporúčame vytvoriť programy na rozvíjanie schopností dlhodobo nezamestnaných pre vyhľadávanie zamestnania. V rámci škôl, podnikov a zamestnaneckých organizácií je potrebné zvýšiť aktivitu a tak predchádzať prerastaniu nezamestnanosti do dlhodobej nezamestnanosti. Je dôležité tiež využívanie rozhlasu, televízie, tlače a propagačného materiálu, ktoré sú bežne dostupné občanom, umožnenie využívania telefónov, počítačov a faxov na pomoc pri hľadaní zamestnania.

4. Záver

Ak zhrnieme uvedené záverečné vyjadrenia môžeme konštatovať, že dosiahnuté výsledky presvedčivo naznačili aplikačné možnosti nami predloženého postupu, že adaptívne prístupy k modelovaniu časových radov viedli k získaniu preukazne kvalitných prepočtov predpovedí vývoje evidovanej miery nezamestnanosti. Aj napriek tomu, že predložená metodológia je výpočtovo náročná, hlavne sofistikované kombinované modely, ktoré sú náročnejšie na splnenie istých predpokladov o vlastnostiach analyzovaných časových radov, získané výsledky naznačili opodstatnenosť ich použitia, hlavne v spojení s výkonnou výpočtovou technikou a kvalitným programovým spracovaním.

5. Literatúra

- [1]ARLT J. 1999. Moderní metody modelování ekonomických časových řad, GRADA Publishing, Praha 1999 ISBN 80-716-539-4.
- [2]CIPRA T.1986. Analýza časových řad s aplikacemi v ekonomii, SNTL/Alfa, Praha 1986 s.100 - 203.
- [3]OBTULOVIČ, P., 2008. Adaptívne prístupy k modelovaniu nezamestnanosti v SR. In Forum Statisticum Slovaccum, roč. IV., 2008, č.4/2008 s 86-92. ISSN 1336-7420.
- [4]<http://www.upsvar.sk> (2011)

Adresa autorov

Peter Obtulovič, doc., Ing., CSc.
Tr. A. Hlinku 2
949 76 Nitra
Peter.Obtulovic@uniag.sk

Zuzana Poláková, Ing. PhD.
Tr. A. Hlinku 2
949 76 Nitra
Zuzana.Polakova@uniag.sk

Vplyv doby mletia a teploty žihania na hodnotu koercivity kompaktovaných vzoriek zisťovaný pomocou štatistických metód

The influence of the time of milling and the temperature of annealing to the value of the coercivity of compacted samples investigated by statistical methods

Denisa Olekšáková, Gabriela Ižaríková

Abstract: The aim of this paper is to investigate the influence of the time of milling and the temperature of annealing of bulk compacted samples NiFe to the value of the coercivity using by the statistical methods. We use the method of least squares and the method of multiple regression.

Key words: coercivity, time of milling, temperature of annealing, compacted samples

Kľúčové slová: koercivita, doba mletia, teplota žihania, kompaktované vzorky

JEL classification: C15, C20, C30

1. Úvod

Daný príspevok je zameraný na analýzu magnetických vlastností, hlavne koercivity, kompaktovaných vzoriek NiFe (81 hm.% Ni, 19 hm.% Fe). Kompaktované vzorky vo forme malého valčeka (priemer 10 mm, výška 3 mm, váha 2 g) sme pripravili stláčaním prášku v lisovacom zariadení (pri tlaku 800 MPa a pri teplote lisovania 600 °C). Prášok sme získali mletím tenkej pásky rovnakého chemického zloženia v planetárnom mlyne po dobu 10, 15, 20, 25 a 30 hodín. Pri každej tejto dobe mletia sme mlyn otvorili a odobrali potrebné množstvo prášku na kompaktovanie vzorky. Pripravili sme tak sériu vzoriek, ktoré sme ďalej postupne žihali pri teplote 500 °C a 1000 °C a tak sledovali ich magnetické vlastnosti, hodnoty koercivity, v závislosti od doby mletia a teploty žihania [1].

2. Koercivita

Koercivita sa tiež nazýva koercívna sila. Je to schopnosť permanentného magnetu odolávať demagnetizácii externým magnetickým poľom a tiež vlastným demagnetizačným poľom. Existujú dva typy koercivity. Tzv. "skutočná" koercivita, jednoducho nazývaná koercivita, znamená magnetické pole, pri ktorom je celková indukcia v magnetе nulová. Druhý typ tzv. "vnútorná" koercivita znamená pole, pri ktorom je celková polarizácia nulová (vektory polarizácie individuálnych magnetických domén sa vzájomne rušia) [2].

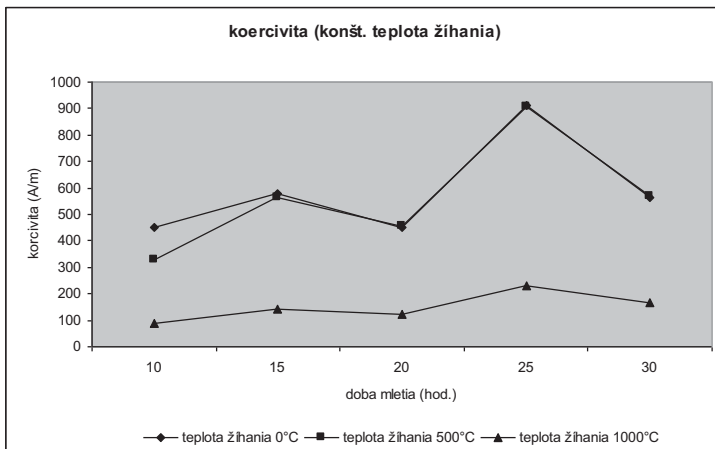
3. Analýza závislosti hodnoty koercivity od doby mletia a od teploty žihania

Tabuľka č. 1: Hodnoty koercivity pre vzorku žihanú pri teplote (0 °C, 500 °C, 1000 °C) a vzorku pripravenú z prášku mletého 10, 15, 20, 25 a 30 hod.

doba mletia (hod.)	koercivita (A/m) (teplota žihania 0°C)	koercivita (A/m) (teplota žihania 500°C)	koercivita (A/m) (teplota žihania 1000°C)
10	452	329	88
15	578	564	141
20	453	456	124
25	912	905	232
30	563	568	165

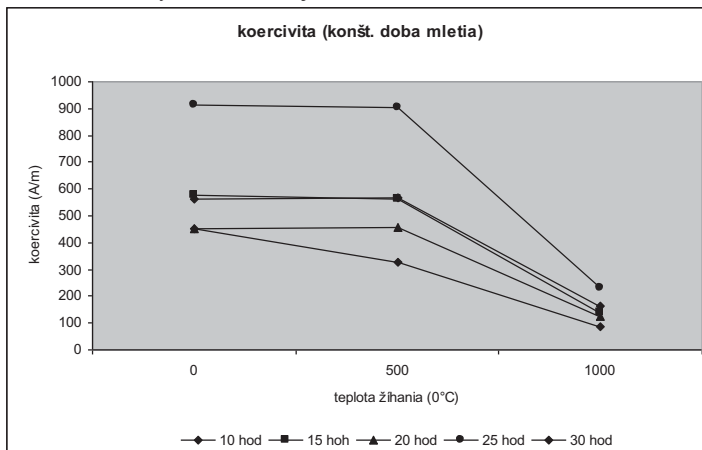
V tabuľke sú údaje o hodnotách koercivity pre vzorku žihanú pri rôznych teplotách (0 °C, 500 °C, 1000 °C) a rôznych dobách mletia (10, 15, 20, 25, 30 hod.). Pri skúmaní

koercivity v závislosti od doby mletia a od teploty žihania je možné preskúmať zmenu koercivity spôsobenú len zmenou doby mletia, zmenu koercivity spôsobenú len zmenou teploty žihania a zmenu koercivity spôsobenú obidvoma faktormi, čiže aj dobou mletia aj teplotou žihania.



Graf č. 1: Porovnanie hodnôt koercivity pri konštantnej teplote žihania v závislosti od doby mletia.

Pri analýze hodnôt podľa grafu č.1, kde sú znázornené hodnoty koercivity pri konštantnej teplote žihania je zrejme, že doba mletia nemení jej hodnoty jednoznačne. Hodnota koercivity pre vzorku žihanú pri 1000 °C vykazuje rast, ale pri teplotách 0 °C a 500 °C to nemožno povedať, hodnoty striedavo rastú a klesajú. Z toho vyplýva, že nemožno tvrdiť, že s nárastom doby mletia rastie aj koercivita.



Graf č. 2: Porovnanie hodnôt koercivity pri konštantnej dobe mletia v závislosti od teploty žihania.

Pri znázornení hodnôt koercivity v závislosti od zmeny teploty, ak považujeme dobu mletia za konštantnú, vyplýva, že pri zmene teploty žihania z 0°C na 500°C sa koercivita nemení, ale pri zmene teploty žihania z 500°C na 1000°C koercivita prudko klesla pri každej dobe mletia (graf č. 2).

Z týchto analýz vyplýva, že nemožno jednoznačne pripísať zmenu koercivity len zmene doby mletia resp. zmene teploty žihania, tzn. že nezávisí len od jednej premennej.

Koercivita (Hc) závisí od doby mletia – premenná X a od teploty žihania – premenná Y . Túto závislosť je možné popísať vzťahom:

$$Hc = a_0 + a_1x + a_2y,$$

kde a_0, a_1, a_2 sú neznáme parametre. Pre odhad parametrov pomocou metódy najmenších štvorcov platí [3]:

$$S(a_0, a_1, a_2) = \sum_{i=1}^n (Hc_i - a_0 - a_1x_i - a_2y_i)^2 \text{ (min. hodnota súčtu štvorcov odchýlok).}$$

Parametre a_0, a_1, a_2 sú riešením nasledujúcej sústavy rovníc:

$$\begin{aligned} a_0n + a_1\sum_{i=1}^n x_i + a_2\sum_{i=1}^n y_i &= \sum_{i=1}^n Hc_i \\ a_0\sum_{i=1}^n x_i + a_1\sum_{i=1}^n (x_i)^2 + a_2\sum_{i=1}^n x_i y_i &= \sum_{i=1}^n x_i Hc_i \\ a_0\sum_{i=1}^n y_i + a_1\sum_{i=1}^n x_i y_i + a_2\sum_{i=1}^n (y_i)^2 &= \sum_{i=1}^n y_i Hc_i \end{aligned}$$

Parametre a_0, a_1, a_2 je možné určiť aj pomocou matematického softvéru (pri spracovaní bola použitá česká verzia Excelu).

	Koeficienty	Chyba stf. hodnoty	t stat	Hodnota P	Dolní 95%	Horní 95%
Hranice	440,1333	147,8130	2,9776	0,0115	118,0764	762,1902
Soubor X 1	10,8	6,4511	1,6741	0,1199	-3,2557	24,8557
Soubor X 2	-0,4416	0,1117	-3,9522	0,0019	-0,6851	-0,1981

Pre závislosť koercivity od doby mletia (hod.) a teploty žihania (°C) platí:

$$Hc = 440,1333 + 10,8x - 0,4416y.$$

Parameter a_1 je kladný, čiže s rastom doby mletia mierne narastá aj koercivita, parameter a_2 je záporný, z toho vyplýva, že čím je vyššia teplota žihania, tým je nižšia hodnota koercivity. Z toho teda vyplýva, že najideálnejšia je kombinácia krátkej doby mletia pri vysokej teplote žihania.

4. Záver

V predkladanom článku analyzujeme vplyv teploty žihania kompakovaných vzoriek a doby mletia prášku, z ktorého boli dané vzorky pripravené, na hodnoty koercivity vzoriek. Zistili sme, že hodnoty koercivity a teda aj magnetické vlastnosti materiálov, sú ovplyvnené oboma parametrami, dobou mletia prášku aj teplotou žihania vzorky. Zvyšujúcou sa dobou

mletia koercivita narastá, čo je dôsledok štruktúrnych zmien pri lámaní a nasledovnom spájaní sa častíc prášku pri mletí. Pri zvyšovaní teploty žihania dochádza k poklesu koercivity, čo má za následok zníženie vnútorných napätí, ktoré boli do vzorky indukované počas procesu mechanického mletia a lisovania. Zistili sme, že ak chceme získať vzorku s čo najlepšimi magneticky mäkkými vlastnosťami, potrebujeme ju pripraviť z prášku mletého čo najkratšiu dobu a zároveň žihať pri čo najvyššej teplote.

5. Literatúra

[1] OLEKŠÁKOVÁ, D.: Dizertačná práca, Košice, 2007.

[2] <http://sk.wikipedia.org/wiki>

[3] GRINČOVÁ, A., IŽARÍKOVÁ, G., PIRČ, V. : Časové rady, Košice 2009.

Adresa autora:

Mgr. Gabriela Ižaríková, PhD.

Letná 9

040 01 Košice

gabriela.izarikova@tuke.sk

RNDr. Denisa Olekšáková, PhD.

Letná 9

040 01 Košice

denisa.oleksakova@tuke.sk

The analysis of exams outcomes of the mathematical compulsory and optional subjects

Análýza výsledkov skúšok z matematických povinných a voliteľných predmetov

Dana Országhová, Radomíra Gregáňová

Abstract: In the paper we present the analysis of student study results of compulsory and optional mathematics subjects. Our sample of students is from the Faculty of Economics and Management at the Slovak University of Agriculture in Nitra. The content of the optional subject belongs to the application in the field of financial and actuarial mathematics. The incorporation of applied tasks into the mathematics increases interest to study mathematical subjects. Feedback is an important part of educational process and through analysis of test results we can indicate its quality. The main aim of this paper is the statistical analysis of tests results of compulsory subjects Mathematics and optional subject Basics of Insurance Mathematics. By χ^2 – test of data independence we verify the hypothesis if the data dependence exists between the exams outcomes of mentioned subjects.

Key words: teaching of mathematics, financial and actuarial mathematics, statistical methods, investigation of data dependence, statistical characteristics

JEL classification: I 21, C 12

1. Introduction

Education and its quality belong to the frequent topics that are discussed in the context of the transformation of the educational system at all levels – primary, secondary and university level. In addition to qualitative methods, we can evaluate the quality of learning outcomes by the exact methods and tools of mathematical statistics. Interesting objects for investigation in the teaching of mathematics are compulsory and optional subjects which are the main theme of our contribution.

The graduates of the Faculty of Economics and Management, the Slovak University of Agriculture (FEM SUA) in Nitra, are professionally trained to work in the fields of economics, management, marketing, banking, finance, insurance, social services, in general, and in many other areas. Basic knowledge of higher mathematics student obtained by passing compulsory subjects Mathematics I, II, respectively Mathematics A, B in the 1st year of bachelor's degree. These subjects provide a basic course of higher mathematics with examples of applications in economics courses. Financial and actuarial mathematics provides the appropriate application of mathematics in the financial and insurance field. By solving practical problems the mathematical education becomes more attractive for students. Knowledge of financial mathematics allows lending or investing funds more efficient and rational way. Students can apply acquired knowledge and methods in decision in their future job, but also in deciding on the assessment of private finance [2].

Present educational process requires the active participation of students to acquire new knowledge. Teachers stress and apply methods of the individual study, for example it is increasing the usage of electronic study materials. This is very difficult for the students in the 1st year because they have to cope with changes regarding new requirements and a system of studying being moved from the secondary school to the university. In the teaching of mathematical subjects it is important to consider specificity of the process of acquisition and fixation of mathematical knowledge. We are interested, how all these conditions influence the quality of education in the field of mathematical subjects [3].

Study results and outcomes are affected by:

- The adaptation to the university system study,
- Searching and applying the methods of individual study,
- Awareness of own responsibility for learning outcomes,
- Knowledge of secondary school,
- Relationship to the study subject [1].

2. Material and methods

The main motivation to write this paper follows from the question: is there any relationship between students' knowledge and exams of compulsory and optional subjects? Our material source was collected from the teaching of compulsory subjects in the 1st study year Mathematics I, Mathematics II (respectively Mathematics A, B) and of optional subject Basics of Insurance Mathematics in the 2nd or 3rd study year. Both mentioned subjects end with the exam and our methodological approach is based on the analysis of the students study results – exams outcomes. By statistical methods we have tested the hypothesis if the data dependence exists among the assessment results of the compulsory and optional subjects.

We describe used methodology: We have statistical sample of range n and we investigate two statistical attributes – the first one signed by X is the test mark of the subjects Mathematics II, Mathematics B (1st study year), the second one signed by Y is the test mark of the subject Basics of Insurance Mathematics (3rd study year). The primary table for quantitative characteristics is a sequence of ordered pairs of values, which are further sorted out into contingency table of two-dimensional frequency distribution.

Then we have determined these two hypotheses:

- Hypothesis H_0 : Attributes X and Y are independent,
- Alternative hypothesis H_1 : Attributes X and Y are dependent.

- Test criterion (for large n):
$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \left(\frac{n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}}{\frac{n_{i\bullet} n_{\bullet j}}{n}} \right)^2,$$

where n_{ij} is empirical frequency, $\frac{n_{i\bullet} n_{\bullet j}}{n}$ is theoretical frequency.

- Critical region: $W_\alpha = (\chi^2((r-1)(s-1)); \infty)$,

where $\chi^2((r-1)(s-1))$ is critical value of χ^2 – probability distribution.

True: If $\chi^2 \in W_\alpha$, so the hypothesis H_0 is rejected in favor of the alternative hypothesis;

if $\chi^2 \notin W_\alpha$, so the hypothesis H_0 is not rejected.

To preserve the objectivity of the dependency analysis and demonstrativeness of used statistical methods, we explored the classification results of academic year from 2003/2004 till 2009/2010. In these years test results have been consistently rated by scale A (1), B (1.5), C (2), D (2.5), E (3) and FX (4). The option of the final evaluation FX (4) did not occurred in our statistical sample.

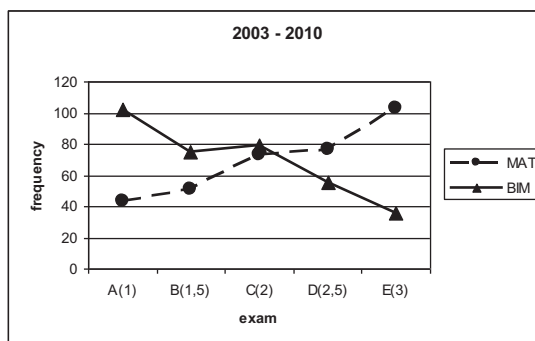
We realized the selection by single survey, non-anonymously, and then we registered marks in the table with the usage of MS Excel 2003. Then the list has been changed to anonymous (students appear in the tables just below the serial number). We made data processing of the statistical sample by statistical methods, namely, we tested the hypothesis of the distribution function using the χ^2 – test of independence [4], [5], [6].

3. Results and discussion

Table 1: Data arranged in contingency table

MAT \ BIM	1	1,5	2	2,5	3	Total
1	26	8	4	5	0	43
1,5	24	16	9	1	1	51
2	27	14	18	10	4	73
2,5	19	20	19	9	10	77
3	6	17	29	30	21	103
Total	102	75	79	55	36	347

In the Table 1 there are given sorted data: ordered pair $[x, y]$ where x is a classification of Mathematics (abbreviation: MAT) and y is a classification of optional subject Basics of Insurance Mathematics (abbreviation: BIM). The data are also graphically displays (Graph 1).



Graph 1: Study results of subjects MAT and BIM

From the graphic representation of exam outcomes we indicate principal differences. There are several factors that affect this state. The compulsory subject is taught in the first study year and it is associated with the students' adaptation to the university system. Students come from different types of secondary schools and have different range of mathematical knowledge. Many of students wrongly assume that the university study will not contain subject mathematics. Moreover, the requirement of the individual study is difficult for some students, and this is reflected in the final exam.

In the Table 2 we can see that in academic year 2006-2007 we have no data to analyze. After the process of faculty accreditation in 2005 new study programs were established and many changes have occurred in optional subjects – many courses were omitted. From the next academic year 2006-2007, student participation on the optional subject Basics of Insurance Mathematics was lower.

The mentioned optional subject was included in the 2nd year of bachelor degree study and after accreditation in the year 2005 was moved to 3rd year. It is clear that students are already adapted to the university study system and achieve better evaluation outcomes of this

subject. The thematic content is not abstract as mathematics, which facilitates its understanding. Students are familiar with some terms of financial and actuarial mathematics from specialized subjects. Another motivation to get better assessment is a study average as a condition for continuation in the second – engineer degree study.

In the Table 2 there are listed the results of data analyzing and statistical testing of hypothesis:

- Hypothesis H_0 : Attributes X and Y are independent,
- Alternative hypothesis H_1 : Attributes X and Y are dependent.

Table 2: Analysis of overall results

		Year 03-04	Year 04-05	Year 05-06	Year 07-08	Year 08-09	Year 09-10	Total
Frequency	MAT	75	100	96	13	25	38	347
	BIM	75	100	96	13	25	38	347
Arithmetic mean	MAT	2,386	2,355	2,198	2,154	1,58	1,96	2,21
	BIM	1,773	1,84	1,619	1,577	1,96	1,97	1,78
Correlation coefficient		0,621	0,459	0,533	0,371	0,627	0,355	0,451
Test criterion χ^2		37,57	34,21	42,219	13	28,155	20,99	90,41

In the table we present obtained results of testing hypotheses for each academic year and the data together for the period 2003-2010. We have made calculations using the spreadsheet MS Excel 2003. We found the critical value χ^2 in tabulating values $\chi_{0,05}^2(4 \cdot 4) = 26,30$.

We can state:

- Tested data are dependent in academic year 03-04, 04-05, 05-06, 08-09 and for all academic years 2003-2010 combined because for these data the value of each χ^2 -test is greater than the critical value of 26.30 for the significance level 0,05. Therefore, we can reject the hypothesis H_0 with 95% confidence.
This means that we accept the alternative hypothesis H_1 : Attributes X and Y are dependent.
- Tested data are independent in academic year 07-08 and 09-10. The value of each χ^2 -test is less than the critical value of 26.30 for the significance level 0.05.
Therefore, we cannot reject the hypothesis H_0 , that attributes X and Y are independent.

4. Conclusion

In the paper we have presented one method of testing statistical hypotheses and we used it on data from the teaching process. The educational process has been going through various changes. We have depicted the educational content from the field of financial and insurance mathematics that can be very useful for future economists and managers that are trained at the Faculty of Economics and management. Our experience from the teaching confirms that

students are interested in optional subjects that are oriented on real application of mathematical methods.

Compulsory and optional mathematical subjects are concentrated in the bachelor degree study. Optional subject Basics of Insurance Mathematics includes the main topics of financial and actuarial mathematics. The knowledge acquired from this course are useful for real life, for example, when we are deciding how to obtain or invest money, or use some of financial products: loans, mortgages, savings, building savings, construction loans, retirement savings and so on. When students choose this subject from offering of study program, we assume that they are interested in studying its contents. Besides the theoretical part contains the financial mathematics also computing, which students can realize through software products, such as MS Excel. And this will also support the development of cross-curricular activities with other specialized subjects.

Knowledge testing and assessment is the important part of the educational process. From obtained data and their analysis it is evident that now we can be relative satisfied with them. Our current task would be – to keep students' interest in this optional subject.

5. References

- [1]GREGÁĽOVÁ, R. – ORSZÁGHOVÁ, D. 2010. Applied problems of Financial Mathematics on the WEB Sites. In 11th International Conference Global Economy: Chalenges and Perspectives (International Scientific Days 2010) – Proceedings of International Scientific Conference (CD-ROM). Nitra, 2010, p. 2501-2524. ISBN 978-80-552-0385-0
- [2]ORSZÁGHOVÁ, D. 2009. Použitie matematických metód vo finančnom sektore. In Zborník príspevkov zo 6. žilinskej didaktickej konferencie s medzinárodnou účasťou DidZA 2009 (CD nosič). Žilina: Žilinská univerzita, 2009, 5 s. ISBN 978-80-554-0050-1
- [3]ORSZÁGHOVÁ, D. – GREGÁĽOVÁ, R. – MAJOROVÁ, M. 2007. Professional Training of Economists and Managers in the Context of the Transformation of University Education. In: Scientific papers (CD) “The Path of Internationalization and Integration in the Europe of Regions”. Curtea de Arges, Romania, 2007, p. 259-264. ISBN 978-80-8069-857-7
- [4]TIRPÁKOVÁ, A. – MALÁ, D. 2007. Základy štatistiky (Basics of Statistics). Nitra, UKF, 2007, 166 p. ISBN 978-80-8094-220-5
- [5]VRÁBELOVÁ, M. – MARKECHOVÁ, D. 2001. Probability and Statistics (Pravdepodobnosť a štatistika). Nitra, UKF, 2001, 199 p. ISBN 80-8050-429-6
- [6]WIMMER, G. 1993. Statistical Methods in Education (Štatistické metódy v pedagogike). Hradec Králové, GAUDEAMUS, 1993, 1st edition, 154 p. ISBN 80-7041-864-8

Acknowledgement

The paper was supported by grant from Grant Agency KEGA: “Theoretical and mathematical transformation of the educational training of agricultural engineers“, No. 3/7382/09.

Adresa autorov:

Dana Országhová, doc., RNDr., CSc.,
Katedra matematiky FEM SPU Nitra
Tr. A. Hlinku 2
949 76 Nitra
Dana.Orszaghova@fem.uniag.sk

Radomíra Gregáňová, Mgr., PhD.,
Katedra matematiky FEM SPU Nitra
Tr. A. Hlinku 2
949 76 Nitra
Radomira.Greganova@fem.uniag.sk

Prognóza vývoja priemerných cien bytov v Slovenskej republike Prognosis of the average apartment prices in Slovakia

Miroslav Pánik

Abstract: This article describes a method of forecasting average prices of apartments in the Slovak Republic. For creating a forecast the neural network is used. Ex-post forecast showed that the modeled data satisfactorily describe the actual value of the average prices of flats. Model can capture the trend in price development. Subsequently, the model is applied and the result is ex-ante forecasts for three periods ahead.

Key words: real estate, development of house prices, forecast, neural networks

Kľúčové slová: realitný trh, vývoj cien bytov, prognóza, neurónové siete,

JEL classification: C45

1. Úvod

Sledovanie vývoja cien nehnuteľností sa stalo veľmi dôležité, najmä po prepuknutí hospodárskej krízy, nakoľko jednou z hlavných príčin bol neprimeraný rast cien nehnuteľností v USA, čo viedlo k realitnej bubline a turbulenciám na svetových finančných trhoch. Ceny nehnuteľností sú ovplyvňované množstvom dopytových a ponukových faktorov, ktoré pôsobia na realitnom trhu.¹ Ceny nehnuteľností na Slovensku dosahovali maximálne hodnoty v prvej polovici roku 2008, následne od tretieho štvrťroka 2008 je pozorovaný prepad cien.

Cieľom príspevku je prognóza vývoja priemerných cien bytov v SR na tri obdobia dopredu.

2. Metódy a údajová základňa

Údajová základňa priemerných cien bytov v SR je čerpaná z databáz Národnej banky Slovenska². Štvrťročné údaje sú dostupné od prvého štvrťroka 2005.

Aplikácia štandardných ekonometrických postupov poskytuje rozporuplné hodnoty odhadnutých parametrov.³ Podrobná analýza je dostupná v [1]. Na vytvorenie prognózy vývoja priemerných cien bytov bola použitá neurónová sieť. Zostrojovanie neurónovej siete bolo realizované pomocou softvéru Statistica, modul Neural Networks.

Pri vyberaní vhodnej neurónovej siete boli porovnávané koeficienty korelácie a chyby jednotlivých modelov. Najvýhodnejšia zo všetkých porovnávaných bola sieť RBF (Radial basis function) s jednou skrytou vrstvou s piatimi neurónmi. RBF sieť je druh neurónovej siete, kde sa ako prechodová funkcia používa radiálna bázová funkcia. Radiálna funkcia je špeciálna matematická funkcia, ktorá so zväčšujúcou sa vzdialenosťou od centra monotónne klesá resp. stúpa. Jednovrstvovú RBF sieť môžeme zapísať ako:

$$f(x) = w_0 + \sum_{i=1}^l w_i p_i(x) \quad (1)$$

Kde w_i – váhy

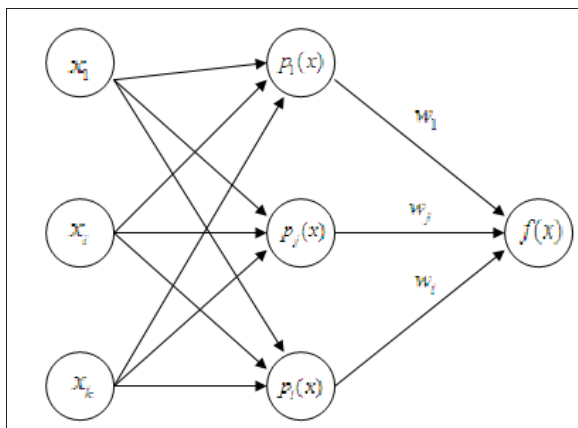
p_i – prechodové funkcie.

¹ Cár, M. 2009.: Výber faktorov ovplyvňujúcich ceny nehnuteľností na bývanie na Slovensku; Biatec č.3/2009

² <http://www.nbs.sk/sk/statisticke-udaje/vybrane-makroekonomicke-ukazovatele/ceny-nehnutelnosti-na-byvanie>

³ Karšay, A. 2010.: Prognózy vývoja cien nehnuteľností na bývanie v prostredí krátkych časových radov
http://www.nbs.sk/_img/Documents/PUBLIK/MU/ceny_nehnut_prog.pdf

Všetky hodnoty z k vstupujúcich hodnôt vektora x sa použijú ako parameter pre prechodové funkcie $p_i, \forall i \in K: 0 < i \leq l$. Výstupom siete je $f(x)$, lineárna kombinácia prechodových funkcií a váh. Na obrázku nižšie je zobrazená RBF sieť s jednou skrytou vrstvou.



Obrázok 1: RBF sieť

Na prognózu cien bytov v SR bola použitá RBF neurónová sieť s jednou skrytou vrstvou s piatimi neurónmi. Zvolená sieť vykazuje najvyššiu koreláciu (0,982) a najmenšie chyby spomedzi všetkých porovnávaných modelov. Celkovo bolo 24 pozorovaní rozdelených v pomere 2:1:1, do troch skupín s nasledujúcou početnosťou. Tréningové údaje: 12, verifikačné údaje: 6, testovacie údaje: 6. Ďalšie charakteristiky sú v nasledujúcich tabuľkách.

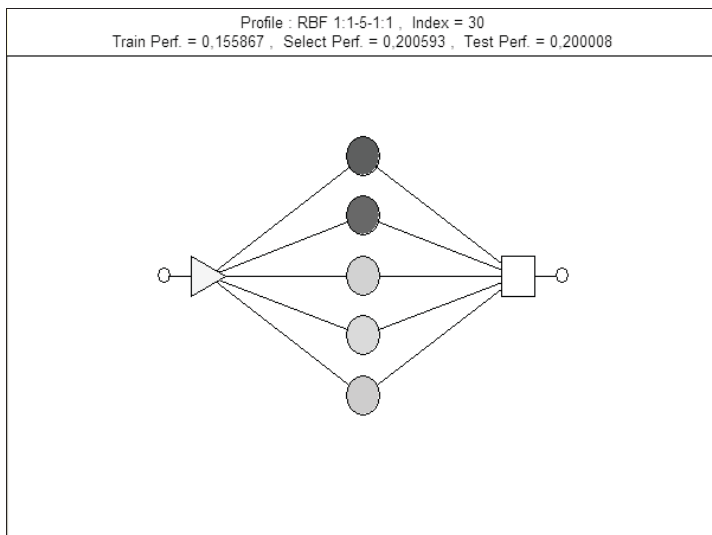
3. Výsledky

Tabuľka 1: Chyby najmenších štvorcov pre jednotlivé skupiny údajov

Charakteristika	Tréningové údaje	Verifikačné údaje	Testovacie údaje
RMS error	0,000549	0,000696	0,000851

Tabuľka 2: Charakteristiky regresie

Charakteristika	Tréningové údaje	Verifikačné údaje	Testovacie údaje
Data Mean	1250,333	1277,167	1153,333
Data S.D.	260,078	253,270	214,230
Error Mean	0,000	7,712	45,931
Error S.D.	40,537	50,804	42,848
Abs E. Mean	30,871	32,098	50,057
S.D. Ratio	0,156	0,201	0,200
Correlation	0,988	0,980	0,981



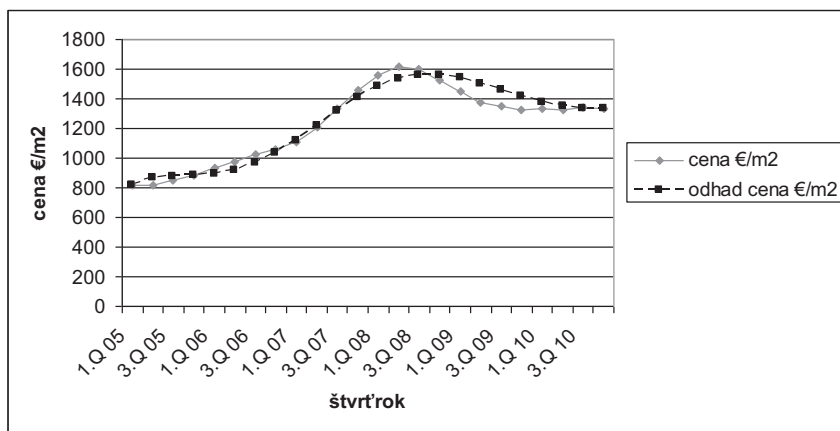
Obrázok 2: Neurónová sieť RBF s jednou skrytou vrstvou s piatimi neurónmi

Pomocou neurónovej siete je realizovaná prognóza. Prognóza ex-post ukázala, že namodelované dáta uspokojivo vystihujú skutočné napozorované hodnoty priemerných cien bytov. Model dokáže zachytiť trend vývoja cien. Najmenšia odchýlka je v štvrtom štvrťroku 2005 a to -0,684, kedy aproximovaná hodnota je 880,316 a skutočná napozorovaná hodnota je 881. Naopak najväčšiu odchýlku pozorujeme v druhom štvrťroku 2009, kedy odchýlka je 131,906. Skutočné a modelované hodnoty sú v nasledujúcej tabuľke a grafe.

Tabuľka 3: Modelované a skutočné priemerné ceny bytov v SR

štvrťrok	cena €/m ²	odhad cena €/m ²	rezidual
1.Q 05	815	819,542	4,542
2.Q 05	814	863,082	49,082
3.Q 05	852	876,495	24,495
4.Q 05	881	880,316	-0,684
1.Q 06	933	890,205	-42,795
2.Q 06	972	916,497	-55,503
3.Q 06	1 029	964,186	-64,814
4.Q 06	1 059	1033,325	-25,675
1.Q 07	1 108	1119,827	11,827
2.Q 07	1 206	1216,569	10,569
3.Q 07	1 332	1314,710	-17,290
4.Q 07	1 458	1405,073	-52,927
1.Q 08	1 555	1479,464	-75,536
2.Q 08	1 619	1531,783	-87,217
3.Q 08	1 600	1558,810	-41,190
4.Q 08	1 521	1560,581	39,581
1.Q 09	1 447	1540,318	93,318

2.Q 09	1 372	1503,906	131,906
3.Q 09	1 354	1458,977	104,977
4.Q 09	1 327	1413,697	86,697
1.Q 10	1 332	1375,381	43,381
2.Q 10	1 329	1349,079	20,079
3.Q 10	1 340	1336,288	-3,712
4.Q 10	1 332	1333,919	1,919



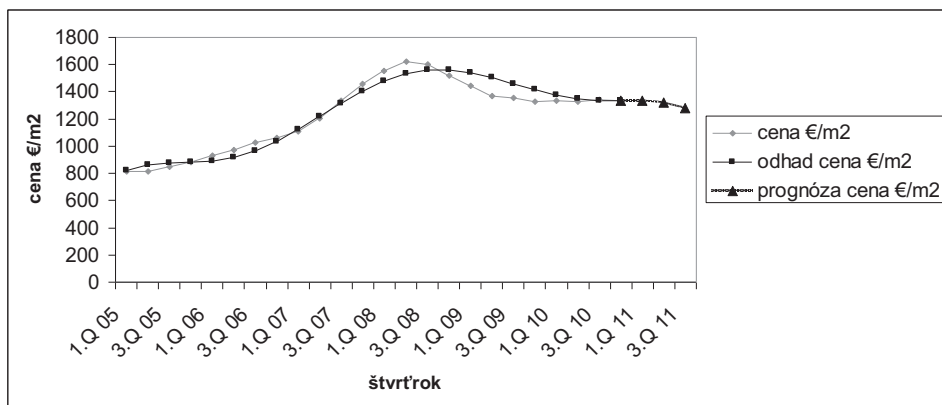
Obrázok 3: Modelované a skutočné ceny bytov

Prognóza ex-ante na tri štvrťroky dopredu ukazuje stagnáciu cien v prvom štvrťroku 2011 a následný pokles cien v druhom a treťom štvrťroku. Výsledná prognóza je zobrazená v tabuľke a grafe.

Tabuľka 4: Prognóza cien bytov

štvrťrok	cena €/m ²	odhad cena €/m ²
1.Q 05	815	819,542
2.Q 05	814	863,082
3.Q 05	852	876,495
4.Q 05	881	880,316
1.Q 06	933	890,205
2.Q 06	972	916,497
3.Q 06	1 029	964,186
4.Q 06	1 059	1033,325
1.Q 07	1 108	1119,827
2.Q 07	1 206	1216,569
3.Q 07	1 332	1314,710
4.Q 07	1 458	1405,073
1.Q 08	1 555	1479,464
2.Q 08	1 619	1531,783
3.Q 08	1 600	1558,810

4.Q 08	1 521	1560,581
1.Q 09	1 447	1540,318
2.Q 09	1 372	1503,906
3.Q 09	1 354	1458,977
4.Q 09	1 327	1413,697
1.Q 10	1 332	1375,381
2.Q 10	1 329	1349,079
3.Q 10	1 340	1336,288
4.Q 10	1 332	1333,919
1.Q 11		1333,636
2.Q 11		1321,633
3.Q 11		1278,877



Obrázok 4: Prognózované a skutočné ceny bytov

4. Záver

Modelované hodnoty sa príliš neodlišujú od skutočných napozorovaných hodnôt cien bytov. Prognóza vytvorená na tri obdobia dopredu ukazuje mierne klesajúci trend vývoja cien bytov. Pri použití metódy neurónových sietí nie je potrebné odhadnúť tvar regresnej krivky a taktiež je odstránený problém nelinearity. Výhodou neurónových sietí je tiež možnosť efektívneho prognózovania aj v prípade krátkeho časového radu. Nedostatkom môže byť menšia efektívnosť dlhodobějších prognóz.

5. Literatúra

- [1] KARŠAY, A. 2010.: Prognóza vývoja cien nehnuteľností na bývanie v prostredí krátkych časových radov; http://www.nbs.sk/_img/Documents/PUBLIK/MU/ceny_nehnut_prog.pdf
- [2] CÁR, M. 2009.: Výber faktorov ovplyvňujúcich ceny nehnuteľností na bývanie na Slovensku; Biatick č. 3/2009
- [3] KŘIVAN, M. 2008.: Úvod do umělých neuronových sítí. Vysoká škola ekonomická v Praze, Nakladatelství Oeconomica. 2008, ISBN 978-80-245-1321-8
- [4] BROOMHEAD, S., LOWE, D. 1988.: Multivariate functional interpolation and adaptive networks, Complex systems, 2:321-355

[5] Národná banka Slovenska: <http://www.nbs.sk/sk/statisticke-udaje/vybrane-makroekonomicke-ukazovatele/ceny-nehnutelnosti-na-byvanie>

Adresa autora:

Miroslav Pánik, Ing.

Ústav manažmentu, Slovenská technická univerzita v Bratislave

Vazovova 5

miroslav_panic@stuba.sk

Článok bol financovaný z grantov VEGA č. 1/0613/10 „Nové prístupy k riadeniu využitia územia v kontexte prehľbujúcej sa integrácie Slovenskej republiky v Európskej únii“ a VEGA č. 1/0456/09 „Strategické riadenie nehnuteľností v podmienkach globalizácie“.

Konfidenčný interval pre rozdiel parametrov dvoch nezávislých binomických rozdelení s aplikáciou v klinických pokusoch

The confidence interval for the difference of two independent binomial proportions with application in the clinical trials

Ivana Pobočíková, Radovan Wiszt

Abstract: The confidence intervals for the difference of two independent binomial proportions are often used in clinical trials to compare a new treatment with a standard treatment or placebo. It is known that exact confidence intervals guarantee the coverage probability above or equal to the nominal level and are reliable in situations when the sample sizes are small and the proportions are near to the boundaries 0 or 1. In this paper we illustrate the application of the confidence intervals in the clinical trials. We use real examples from the clinical trials. We consider the exact Chan-Zhang interval, the exact Agresti-Min interval and the asymptotic Wald interval.

Key words: binomial distribution, confidence interval, difference of two proportion, coverage probability, interval length, Chan-Zhang interval, Agresti-Min interval, Wald interval

Kľúčové slová: binomické rozdelenie, konfidenčný interval, rozdiel parametrov, pravdepodobnosť pokrytia, dĺžka intervalu, Chan-Zhangov interval, Agrestiho-Minov interval, Waldov interval

1. Introduction

The confidence intervals for the difference of two independent binomial proportions are often used in clinical trials to compare a new treatment with a standard treatment or placebo. We consider a clinical trial in which we want to compare the efficacy of a new treatment with a standard treatment or placebo. This situation we represent with a 2×2 contingency table

Table 1.

	New treatment	Standard treatment (placebo)
Number of successes	X	Y
Number of failures	$n_1 - X$	$n_2 - Y$
Total	n_1	n_2

where $X \sim Bi(n_1, \pi_1)$ and $Y \sim Bi(n_2, \pi_2)$ are two independent binomial random variables.

Let $\delta = \pi_1 - \pi_2$ is the difference of two binomial proportions, $-1 < \delta < 1$. Let $\pi = \pi_1$ and substitute $\pi_2 = \delta - \pi_1$. The joint probability mass function can be expressed as

$$P(X = x, Y = y) = \binom{n_1}{x} \binom{n_2}{y} \pi^x (1 - \pi)^{n_1 - x} (\pi - \delta)^y (1 - \pi + \delta)^{n_2 - y}, \quad (1)$$

for $x = 0, 1, \dots, n_1$, $y = 0, 1, \dots, n_2$, $\pi_1, \pi_2 \in (0, 1)$, $n_1, n_2 \in \mathbb{N}$. For any given δ the domain of π is

$$D(\delta) = \{ \pi : \max \{ 0, \delta \} \leq \pi \leq \min \{ 1, 1 + \delta \} \}. \quad (2)$$

It is known that π is a nuisance parameter for the inference on δ .

The maximum likelihood estimators (MLE) for π_1 and π_2 from a samples are $p_1 = \frac{X}{n_1}$ and $p_2 = \frac{Y}{n_2}$, respectively, where X is a number of successes in a random sample of size n_1 and Y is a number of successes in a random sample of size n_2 .

To evaluate the treatments difference we want to find the $100 \times (1 - \alpha)\%$ two-sided confidence interval $\langle \delta_L, \delta_U \rangle$ for the difference of two independent binomial proportions. $(1 - \alpha)$ is the desired confidence level, $\alpha \in (0, 1)$.

The literature contains several methods for constructing confidence intervals for the difference of two independent binomial proportions. The asymptotic methods are based on standard normal approximation. The Wald interval is the most common used. The $100 \times (1 - \alpha)\%$ Wald interval is defined by

$$\begin{aligned} \delta_L &= (p_1 - p_2) - k_{1-\frac{\alpha}{2}} \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}, \\ \delta_U &= (p_1 - p_2) + k_{1-\frac{\alpha}{2}} \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}, \end{aligned} \quad (3)$$

where k_α is the α -quantile of standard normal distribution. Wald interval is simple to compute, but it is known that interval performs poorly (Newcombe, 1998). The exact methods are derived from exact binomial distribution. In clinical trials such situations are often when the sample sizes are small or moderate and strict conservatism is required. It is known, that the exact methods guarantee the coverage probability above or equal to the nominal level $(1 - \alpha)$ and are more reliable when the sample sizes are small or when the proportions π_1 or π_2 are near to the boundaries 0 or 1.

In this paper we illustrate the application of the confidence intervals for the difference of two independent binomial proportions in the clinical trials. We use the real data from the clinical trials. We consider two exact methods: the Chan-Zhang interval and the Agresti-Min interval.

2. Alternatives of the confidence intervals

The Chan-Zhang interval and the Agresti-Min interval are test based. They are constructed by inverting a hypothesis test under an appropriate alternative hypothesis.

Chan-Zhang interval. Chan and Zhang (1999) proposed exact confidence interval by inverting two one-sided score tests

$$H_0 : \delta = \delta_0 \text{ versus } H_0 : \delta < \delta_0 \text{ and } H_0 : \delta = \delta_0 \text{ versus } H_0 : \delta > \delta_0.$$

Chan and Zhang used for testing the score test statistic

$$Z(X, Y, \delta_0) = \frac{(p_1 - p_2 - \delta_0)}{\sqrt{\hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2}}, \quad (4)$$

where p_1, p_2 are MLE of π_1, π_2 and \hat{p}_1, \hat{p}_2 are maximum likelihood estimators of π_1, π_2 under the restriction that $\hat{p}_1 - \hat{p}_2 = \delta_0$. Miettinen and Nurminen (1985) showed that \hat{p}_1, \hat{p}_2 can be obtained uniquely by closed form.

Chan and Zhang used maximization method to eliminate the effect of the nuisance parameter π . The exact p -value is maximizing over all possible values of the nuisance parameters π . For given $X = x, Y = y$ are the exact one-sided p -values for δ_0 defined by

$$\beta_{CZL}(x, y | Z, \delta_0) = \max_{\pi \in D(\delta_0)} \left\{ \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} P(X=i, Y=j | \delta_0, \pi) I_1(Z(i, j, \delta_0) \geq Z(x, y, \delta_0)) \right\}, \quad (5)$$

$$\beta_{CZU}(x, y | Z, \delta_0) = \max_{\pi \in D(\delta_0)} \left\{ \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} P(X=i, Y=j | \delta_0, \pi) I_2(Z(i, j, \delta_0) \leq Z(x, y, \delta_0)) \right\}, \quad (6)$$

where $I_1(A \geq B) = \begin{cases} 1 & \text{if } A \geq B \\ 0 & \text{otherwise} \end{cases}$, $I_2(A \leq B) = \begin{cases} 1 & \text{if } A \leq B \\ 0 & \text{otherwise} \end{cases}$ are indicator functions.

The $100 \times (1 - \alpha)\%$ Chan-Zhang interval is defined by

$$\begin{aligned} \delta_L &= \inf_{\delta} \{ \delta; \beta_{CZL}(x, y | Z, \delta_0) > \alpha / 2 \}, \\ \delta_U &= \sup_{\delta} \{ \delta; \beta_{CZU}(x, y | Z, \delta_0) > \alpha / 2 \}. \end{aligned} \quad (7)$$

Agresti-Min interval. This interval is derived similarly to the Chan-Zhang interval. Agresti and Min (2001) proposed exact confidence interval by inverting one two-sided score test

$$H_0 : \delta = \delta_0 \text{ versus } H_0 : \delta \neq \delta_0.$$

For given $X = x, Y = y$ is the exact two-sided p -value for δ_0 defined by

$$\beta_{AM}(x, y | Z, \delta_0) = \max_{\pi \in D(\delta_0)} \left\{ \sum_{i=0}^{n_1} \sum_{j=0}^{n_2} P(X=i, Y=j | \delta_0, \pi) I(|Z(i, j, \delta_0)| \geq |Z(x, y, \delta_0)|) \right\}, \quad (8)$$

where $I(|A| \geq |B|) = \begin{cases} 1 & \text{if } |A| \geq |B| \\ 0 & \text{otherwise} \end{cases}$ is an indicator function. The $100 \times (1 - \alpha)\%$ Agresti-Min interval is defined by

$$\begin{aligned} \delta_L &= \inf_{\delta} \{ \delta; \beta_{AM}(x, y | Z, \delta_0) > \alpha \}, \\ \delta_U &= \sup_{\delta} \{ \delta; \beta_{AM}(x, y | Z, \delta_0) > \alpha \}. \end{aligned} \quad (9)$$

3. Methods used to compare of the performance

The performance of the confidence intervals can be evaluated by the coverage probability, conservatism and the interval length.

The coverage probability of the confidence interval $\langle \delta_L, \delta_U \rangle$ is for fixed $n_1, n_2 \in N$ and $\pi_1, \pi_2 \in (0, 1)$ defined by

$$C_{n_1, n_2}(\pi_1, \pi_2) = \sum_{x=0}^{n_1} \sum_{y=0}^{n_2} \binom{n_1}{x} \binom{n_2}{y} \pi_1^x (1 - \pi_1)^{n_1 - x} \pi_2^y (1 - \pi_2)^{n_2 - y} I(x, y, \pi_1, \pi_2), \quad (10)$$

where $I(x, y, \pi_1, \pi_2) = \begin{cases} 1 & \text{if } \langle \delta_L(x, y), \delta_U(x, y) \rangle \\ 0 & \text{otherwise} \end{cases}$ is an indicator function.

The confidence interval is **strict conservative**, if $C_{n_1, n_2}(\pi_1, \pi_2) \geq 1 - \alpha$ for all $n_1, n_2 \in N$ and for all $\pi_1, \pi_2 \in (0, 1)$.

The **expected length** of the confidence interval is defined by

$$EL_{n_1, n_2}(\pi_1, \pi_2) = \sum_{x=0}^{n_1} \sum_{y=0}^{n_2} [\delta_U(x, y) - \delta_L(x, y)] \binom{n_1}{x} \binom{n_2}{y} \pi_1^x (1 - \pi_1)^{n_1 - x} \pi_2^y (1 - \pi_2)^{n_2 - y}, \quad (11)$$

where $\delta_L(x, y), \delta_U(x, y)$ are bounds of a particular confidence interval.

The better confidence interval is such interval, which coverage probability is close to the nominal level $(1 - \alpha)$. The shorter interval is preferred.

The Chan-Zhang interval and the Agresti-Min interval are strict conservative intervals. These intervals are recommended to be used in clinical trials, when the strict conservatism is required and when the sample sizes are small or moderate. The Agresti-Min interval has coverage probability closer to the nominal level, is less conservative. In some cases the Agresti-Min interval become wider, but generally is shorter than the Chan-Zhang interval.

4. Illustrative examples

Example 1. This data was reported by Harmon et al. (2000). Nausea and vomiting are important side effects both during and after spinal anaesthesia for Caesarian section. The efficacy of acupressure at the P6 point on the right forearm in the prevention of nausea and vomiting during Caesarian section was studied. The randomized double-blind controlled study was designed. A total of 94 patients were randomized into two groups to receive either acupressure or control. The data can be represented in 2×2 table as follows

Table 2. Incidence nausea and vomiting during operations

	Acupressure group	Control group
No	36	22
Yes	11	25
Total	47	47

Let π_1 be the probability that patient has nausea or vomiting during the operation in acupressure group and let π_2 be the probability that patient has nausea or vomiting during the operation in control group. We are interested that both methods are equally effective in prevention of nausea or vomiting during the operation. We obtain the 95% confidence intervals for the treatments difference $\delta = \pi_1 - \pi_2$ (acupressure-control). The MLE of proportions are $p_1 = \frac{11}{47}$ in acupressure group and $p_2 = \frac{25}{47}$ in control group, respectively.

The observed difference $\delta = \pi_1 - \pi_2$ is $-0,2979$, which indicates that the acupressure has noticeable improvement in reduce of nausea and vomiting during operation. The Chan-Zhang interval is $\langle -0,4792; -0,0874 \rangle$, the Agresti-Min interval is $\langle -0,4741; -0,0953 \rangle$ and the Wald interval is $\langle -0,4850; -0,1108 \rangle$. The Chan-Zhang interval is the widest, the Wald interval is the narrowest. The confidence intervals gives to the same conclusion. There is

a statistically significant difference between two proportions (the confidence intervals does not contain 0). This indicates that there is a significant reduction of nausea or vomiting during the operation in acupressure group compared to control group.

Example 2. This data was reported by Varma et al. (2004). The efficacy of PUVA (artificial UV light source) administration with and without wearing clothes in psoriasis was studied. The UV dose in patients wearing the fabric was increased proportional to the sun protective factor of the fabric. The randomized single-blind comparative clinical trial was designed. A total of 21 patients with psoriasis vulgaris were randomly divided into two groups. The study group received bath PUVA therapy wearing the fabric and the control group without the fabric. The data after 12 PUVA treatments can be represented in 2×2 table as follows

Table 3. The number of patients with remission (treatment success) and treatment failure

	Study group	Control group
Treatment failure	0	0
Treatment success	14	7
Total	14	7

Let π_1 be the probability that treatment failures with PUVA in study group and let π_2 be the probability that treatment failures with PUVA in control group. We are interested that both the treatments are equally effective. We obtain the 95% confidence intervals for the treatments difference $\delta = \pi_1 - \pi_2$ (new treatment-current treatment). The MLE of proportions are $p_1 = \frac{0}{14}$ in study group and $p_2 = \frac{0}{7}$ in control group, respectively. The observed difference $\delta = \pi_1 - \pi_2$ is 0, which indicates that there is not improvement. The Chan-Zhang interval is $\langle -0,4468; 0,2472 \rangle$, the Agresti-Min interval is $\langle -0,3736; 0,2383 \rangle$ and the Wald interval degenerate at single point 0. These methods yield quite different results. The Wald interval gives to the conclusion that if nothing by the trials happend, never can occur. The Wald interval do not provide sensible answers. Other confidence intervals gives to the conclusion that the difference between two proportions is no statistically significant (the confidence intervals contain 0). This indicate that both the treatments are equally effective.

5. Concluding Remarks

The confidence intervals for the difference of two independent binomial proportions are used in clinical trials to compare the efficacy of a new treatment with a standard treatment or placebo. When the strict conservatism is a major criterion the Chan-Zhang interval and the Agresti-Min are a good choice.

6. Literature

- [1] AGRESTI, A., MIN, Y., 2001. On small sample confidence intervals for parameters in discrete distributions. In: Biometrics 57, pp. 963-971.
- [2] CHAN, I. S. F., ZHANG, Z., 1999. Test-based exact confidence intervals for the difference of two binomial proportions. In: Biometrics 55, pp. 1202-1209.

- [3] HARMON, B., RYAN, M, KELLY, A., BOWEN, M. 2000. Acupressure and prevention of nausea and vomiting during and after spinal anesthesia for Caesarian section. In: British Journal of Anesthesia 84 (4), pp. 463-467.
- [4] MIETTINEN, O. S., NURMINEN, M. 1985. Comparative analysis of two rates. In: Statistics in Medicine 4, pp. 213-226.
- [5] POBOČÍKOVÁ, I. 2011. Exact and quasi-exact confidence intervals for the difference of two binomial proportions. In: 10th international conference Aplimat, Bratislava 2011, pp. 1609-1617.
- [6] VARMA, S. ET AL. 2004. Clothes-on PUVA in psoriasis: Single blind randomized comparative trial on 21 patients. In: Indian Journal Dermatol Venerol Leprol Vol. 70, pp. 152-155.
- [7] WILSON, E. B. 1927. Probable inference, the law of succession, and statistical inference. In: Journal of the American Statistical Association 22, pp. 209-212.

Adresa autorov:

Ivana Pobočíková, Mgr.
Katedra aplikovanej matematiky SjF
Žilinská univerzita
Univerzitná 1
010 26 Žilina
ivana.pobocikova@fstroj.uniza.sk

Radovan Wiszt, MUDr.
Transplantačné centrum
Univerzitná nemocnica Martin
Kollárova 2
036 01 Maritn
swiszt@gmail.com

Analyza a komparacia demografického vývoja rómskej populácie vo vybraných regiónoch SR

Analysis and comparison of the demographic development of the Roma population in selected regions of Slovakia

Zuzana Poláková, Peter Obtulovič

Abstract: Roma are an ethnic group that shares a common origin, language, lifestyle, cultural features and characteristics. Under realistic assumptions, living in Slovakia approximately 400 000 Roma. From a demographics point of view characterized by their higher birth rates and higher mortality. The main them of this thesis is an analysis and comparison of the demographics development of the Roma population in selected regions of Slovakia in the period 1996 – 2009. We focus on the structure of Roma population, demographics characteristics, development and time series forecast for 2 years.

Key words: population development, Roma, ethnic group

Kľúčové slová: populačný vývoj, Róm, etnikum

Jel classification: C25

1. Úvod

Slovensko možno označiť za multietnický štát. Okrem Slovákov žije na území Slovenskej republiky aj nie zanedbateľný počet príslušníkov etnických menšín. Podľa reálnych odhadov etnické menšiny tvoria v súčasnosti zhruba 20% z celkového počtu majoritného obyvateľstva. Pod rôznorodú etnickú skladbu obyvateľstva sa podpisala predovšetkým geografická poloha Slovenska v strede Európy a historický vývoj spojený s významnými presunmi obyvateľstva.

Podiel rómskej populácie, sa postupne zvyšuje, takže na jednej strane stúpajú náklady na sociálnu pomoc Rómom, a na druhej strane stúpajú straty z nevyužitia ich potenciálu na trhu práce. V predkladanom príspevku sa zaoberáme analýzou a komparáciou demografického vývoja Rómskej populácie vo vybraných regiónoch Slovenskej republiky v období rokov 1996 – 2009, pričom sa zameriavame na Košický a Bratislavský kraj, analýzu vývoja rómskej populácie a prognózu vývoja na dva roky. Údaje o národnosti sa zisťujú pri sčítaní obyvateľov každých 10 rokov. V období medzi sčítaniami sa spracováva bilancia pohybu obyvateľstva podľa národností, ktorá priamo nadväzuje na výsledky príslušného sčítania. Na základe tejto bilancie je možné každoročne získať informácie o počte a štruktúre obyvateľstva podľa národností.

2. Vývoj počtu Rómskej populácie vo vybraných regiónoch SR

Sčítanie obyvateľstva, domov a bytov sa realizuje v SR približne každých 10 rokov. Okrem iných charakteristík sa zisťuje aj národnosť a materinský jazyk obyvateľstva. Obe tieto charakteristiky možno považovať za kultúrne charakteristiky obyvateľstva, ktoré sú v multietnickom štáte veľmi dôležité a zaujímavé. Pri sčítaní sa národnosťou rozumela príslušnosť obyvateľa k národu alebo k etnickej skupine. Rozhodnutie, ku ktorej etnickej skupine sa občan prihlási, bolo na samotnom občanovi. U detí do 15 rokov zapísali údaje o národnosti rodiča. Pred rokom 1991 sa obyvateľstvo prihlasovalo k národnosti, s ktorou sa cítilo vnútorne späté. Bolo ponechané na rozhodnutí občana, ku ktorej národnosti sa prihlási. Bolo možné prihlásiť sa len k vybraným národnostiam, medzi ktorými rómska národnosť

uvedená nebola. Preto sa Rómovia museli hlásiť k niektorej z oficiálne uznaných národností. V roku 1991 mali občania SR prvýkrát možnosť po šesťdesiatich rokoch prihlásiť sa k rómskemu etniku. Túto možnosť využilo len necelých 76 tisíc Rómov [3].

Pre porovnanie vývoja počtu Rómskej populácie sme vybrali regióny s významnou odlišnosťou v zastúpení menovaného etnika. V Košickom a Bratislavskom kraji sme vychádzali z absolútnych údajov o počte Rómov (tabuľka 1).

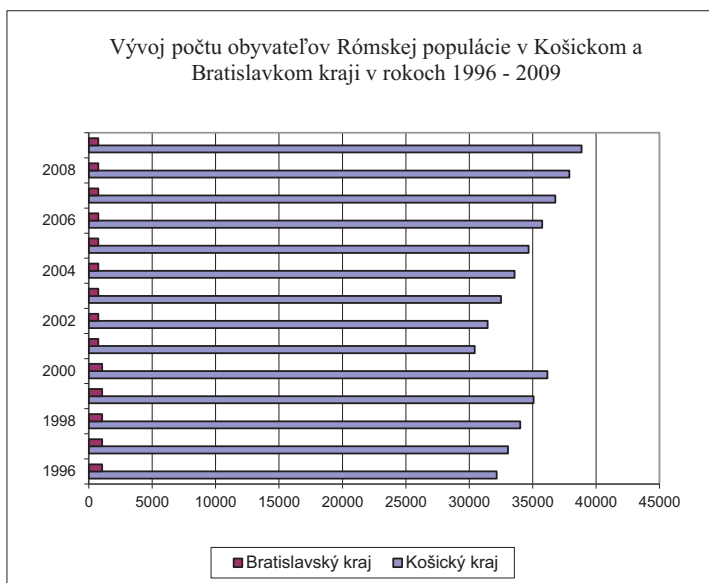
Tabuľka 1 Počet Rómov v Košickom a Bratislavskom kraji v rokoch 1996 až 2009

	1996	1997	1998	1999	2000	2001	2002
KE	32157	33064	34023	35081	36160	30431	31448
BA	1043	1046	1044	1043	1045	753	753
	2003	2004	2005	2006	2007	2008	2009
KE	32510	33567	34682	35748	36785	37894	38859
BA	750	750	750	751	752	750	750

Zdroj: Bilancia pohybu obyvateľstva v Slovenskej republike podľa národnosti, vlastná práca

Z porovnania vybraných regiónov jednoznačne vyplýva (tabuľka1, obrázok 1), že v Košickom kraji žije podstatne viac Rómov ako v Bratislavskom kraji. Vývoj Rómskej populácie v Košickom kraji sa do roku 2000 zvyšoval, no v roku 2001 nastal značný pokles. Po tomto roku badáme každoročný nárast v počte Rómov. Tento výrazný pokles môžeme odôvodniť migráciou Rómov za prácou do Belgicka v tomto období, tiež strachom, ktorého impulzom boli viaceré fakty týkajúce sa diskriminácie Rómov.

Obrázok 1 Vývoj počtu Rómskej populácie v Košickom a Bratislavskom kraji v rokoch 1996-2009



Zdroj: Bilancia pohybu obyvateľstva v Slovenskej republike podľa národnosti, vlastná práca

Najmenej Rómov je v Košickom kraji v roku 2001 (30 431), čo oproti roku 2000 predstavuje výrazný pokles. Najväčší počet Rómov za sledované obdobie je v Košickom kraji v roku 2009 (38 859). Na druhej strane vývoj počtu Rómov v Bratislavskom kraji na začiatku sledovaného obdobia pomerne stagnuje, neskôr mierne klesá, no v roku 2001 bol výrazný pokles (z 1045 na 753). V prípade Košického kraja je situácia v roku 2001 podobná, tento pokles je aj v dôsledku celkovo nízkej natality v rámci celého Slovenska nevynímajúc Rómsku populáciu. Najviac Rómov žilo v Bratislavskom kraji v roku 1997, čo predstavuje 1046 Rómov. Percentuálne zastúpenie Rómskej populácie v Košickom a Bratislavskom kraji z celkového počtu Rómskej populácie v Slovenskej republike je prezentované v tabuľke 2.

Tabuľka 2 Percentuálne zastúpenie Rómskej populácie v Košickom a Bratislavskom kraji z celkového počtu Rómskej populácie v Slovenskej republike

	1996	1997	1998	1999	2000	2001	2002
Košický kraj	37.23	37.67	38.04	38.45	38.84	33.46	33.93
Bratislavský kraj	1.21	1.19	1.17	1.11	0.99	0.83	0.81
	2003	2004	2005	2006	2007	2008	2009
Košický kraj	34.42	34.87	35.33	35.72	36.08	36.43	36.77
Bratislavský kraj	0.79	0.78	0.77	0.75	0.74	0.72	0.71

Zdroj: vlastné výpočty

V košickom kraji je Rómska populácia v roku 1996 zastúpená 37, 23 %, percentuálny rast nastáva do roku 2000. Rápidný pokles Rómskej populácie v Košickom kraji nastal v roku 2001 na úroveň 33,46 %. V roku 2001 značný počet Rómov z Košického kraja vycestovalo za prácou do zahraničia. Situácia na východe Slovenska je z pohľadu nezamestnanosti na poslednom mieste, čo je aj v dôsledku veľkej koncentrácie Rómskej populácie na východ krajiny. V roku 2009 je Rómska populácia zastúpená v Košickom kraji na 36,77 %, táto hodnota je oproti predchádzajúcim ôsmim rokom vyššia v dôsledku zvyšovania natality Rómskej populácie. Bratislavský kraj sa pýši malým percentom zastúpenia Rómov z celkového počtu Rómov na Slovensku. Vývoj zastúpenia Rómskej populácie v Bratislavskom kraji za sledované obdobie má kolísavý priebeh, no s malými zmenami. V roku 1996 žilo v Bratislavskom kraji z celkového počtu Rómov 1,21 %, v roku 2009 bola Rómska populácia v Bratislavskom kraji z celkového počtu Rómov zastúpená 0,71 %.

3. Analýza vývoja Rómskej populácie v Košickom kraji

Jednou z najdôležitejších úloh analýzy časových radov je vystihnúť primeraným spôsobom celkovú tendenciu vo vývoji sledovaného časového radu [2]. Z veľkého okruhu trendových funkcií vyberieme polynomicke funkciu 2. stupňa. Analýzy budú spracované na hladine významnosti $\alpha = 0,05$. Pre analýzu vývoja Rómskej populácie v Košickom sme použili polynomicke trendovú funkciu: variabilita Rómskej populácie v Košickom kraji je týmto regresným modelom vysvetlená na 61,2 %. Regresný model bol zvolený správne (Významnosť $F = 0,005478$) Regresné koeficienty sú štatisticky významné (P hodnoty $b_0 = 1,73E-10$, $b_1 = 0,163252$, $b_2 = 0,038118$). Rovnica krivky má tvar:

$$y = 34520,81x^2 - 712,228x + 72,81044$$

Hodnota priemernej absolútnej percentuálnej chyby (MAPE) pri zvolenom modeli nadobúda hodnotu 0,032227.

V prípade prognózy vývoja rómskeho obyvateľstva je situácia o to komplikovanejšia, že nie sú k dispozícii úplne pravdivé informácie o počte a reprodukčnom správaní Rómov, nakoľko pri sčítaní sa nie každý Róm hlási k rómskemu etniku.

Na základe zvoleného modelu sme realizovali prognózu budúceho vývoja Rómskej populácie v Košickom kraji. Prognózu sme vytvorili pre roky 2010 a 2011. Výsledky prognózy prezentuje tabuľka 3.

Tabuľka 3 Prognóza vývoja Rómskej populácie v Košickom kraji na rok 2010 a 2011

Rok	prognóza
2010	40219
2011	41764

Zdroj: vlastné výpočty

Na základe vykonanej prognózy môžeme konštatovať, že počet Rómov v Košickom kraji v roku 2010 predstavuje pravdepodobné zvýšenie o 1360 Rómov, čo tvorí nárast o 3,5 % oproti predchádzajúcemu roku 2009. V ďalšom roku prognózy by malo dôjsť obdobne ako v predchádzajúcom roku k zvýšeniu o 1545, čo predstavuje nárast o 7,5 % oproti predchádzajúcemu roku. Nárast Rómskej populácie bol v minulosti obmedzený jedným z najstriktejších opatrení, a to zavedením sterilizačného programu pre rómske ženy. V rámci programu sa miestni zdravotnícki pracovníci pokúšali presvedčať rómske ženy, aby sa dali za finančnú kompenzáciu sterilizovať.

4. Analýza vývoja Rómskej populácie v Bratislavskom kraji

Obdobne ako v prvom prípade sme aj pri analýze vývoja Rómskej populácie v Bratislavskom kraji zvolili polynomicke funkciu. Variabilita Rómskej populácie v Bratislavskom kraji je týmto regresným modelom vysvetlená na 86,98 %. Model je vhodný na popis analýzy vývoja (Signifikantné $F = 1,35E-05$). Regresné koeficienty sú štatisticky významné (P hodnoty $b_0 = 6,45E-11$, $b_1 = 0,000268$, $b_2 = 0,00457$). Rovnica krivky má tvar:
 $y = 1186,973x^2 - 78,7034x + 3,441621$

Hodnota priemernej absolútnej percentuálnej chyby (MAPE) pri zvolenom modeli nadobúda hodnotu 0,041088, čo potvrdzuje vhodnosť použitého modelu na prognózu vývoja Rómskej populácie.

Na základe zvoleného modelu sme urobili prognózu vývoja Rómskej populácie v Bratislavskom kraji na ďalšie dva roky (2010, 2011). Výsledky prognózy sú znázornené v tabuľke 4.

Tabuľka 4 Prognóza vývoja Rómskej populácie v Bratislavskom kraji na rok 2010 a 2011

Roky	prognóza
2010	780
2011	808

Zdroj: vlastné výpočty

Na základe prognózy môžeme konštatovať, že tendencia vývoja počtu Rómov v Bratislavskom kraji je rastúca. V roku 2010 bude v Bratislavskom kraji 780 Rómov, oproti predchádzajúcemu roku je to nárast o 4,1 %. Situácia je rovnaká na celom území Slovenska nielen v sledovaných regiónoch. Už vo vyše 60 obciach tvoria Rómovia väčšinu. Predpokladá sa prírastok, že v priebehu desiatich rokov narastie počet rómskych osád o 71. Prírastok rómskeho obyvateľstva je dlhodobo podstatne vyšší v porovnaní s prírastkom majoritnej populácie. Štúdie odhadujú počet Rómov do roku 2025 na približne 500 000, čo predstavuje niekoľkonásobný nárast oproti poslednému sledovanému roku 2009.

5. Záver

Problematika Rómskej populácie je každodennou témou nielen na Slovensku, ale aj v iných krajinách. Počet Rómskej populácie enormne narastá každým rokom. Je zrejmé, že situáciu je potrebné riešiť. Známe je, že časť Rómov sa nehlási k rómskej národnosti kvôli nedostatočnému etnickému vedomiu, ako aj zo strachu z diskriminácie. V príspevku sme sa venovali analyzovaniu a komparácii demografického vývoja Rómskej populácie v Košickom a Bratislavskom kraji. Z analýzy vyplýva, že vývoj počtu Rómskej populácie v Košickom kraji mal rastúci priebeh od roku 1996 - 2000, následne nastal výrazný pokles (5,4 %) v roku 2001. Predpokladáme, že je to v dôsledku migrácie Rómov za prácou do Belgicka, tiež strachom, ktorého nosičom boli viaceré fakty a diskriminačné motívy. Od tohto roku stúpa počet Rómskej populácie až po koniec sledovaného obdobia, kedy tvorí v Košickom kraji 36,77 % z celkového počtu Rómov na Slovensku. Môžeme konštatovať, že Rómovia žijú rozptýlene po celom území Slovenska, no najvyššia koncentrácia je na juhu stredného a východného Slovenska už dlhé desaťročia. Naopak, Rómska populácia v Bratislavskom kraji je zastúpená malým počtom a hoci predpokladáme do budúcnosti nárast, je v tomto regióne minimálna. Rómska populácia v rámci celkovej populácie na Slovensku tvorí v priemere 1,5 – 2 %. Podľa odhadov ŠÚ SR tvoria v skutočnosti Rómovia na Slovensku 8 % z celkovej populácie na Slovensku.

Rozhodujúcim faktorom pre približovanie sa demografických ukazovateľov rómskej a nerómskej časti populácie bude integrovanie sa rómskych občanov do majoritnej spoločnosti. Možno predpokladať, že nízka úroveň pôrodnosti na Slovensku vyvolá reakciu štátu v podobe populačných opatrení, ktoré môžu ovplyvniť aj vývoj rómskej pôrodnosti. Je pravdepodobné, že vývoj rómskej pôrodnosti sa bude znižovať, hoci pomalšie ako u ostatného obyvateľstva.

6. Literatúra

- [1] KLUFOVÁ, R. – POLÁKOVÁ, Z. 2010. Demografické metódy a analýzy. Demografie české a slovenské populace. Wolters Kluwer, 2010. ISBN 978-80-7357-546-5.
- [2] OBTULOVÍČ, P. 2010. Bioštatistika. SPU Nitra 2010. 171 s. ISBN 978-80-552-0397-3.
- [3] POLÁKOVÁ, Z. 2002. Analýza vývoja rómskej populácie v SR a prognóza do roku 2010. Štrukturálny cenzus fariem 2001. SŠaDS. CD
- [4] PRUCHA, J. 2004. Multikultúrní výchova. Praha: ISV, 2004. 223 s. ISBN 80-85866-72-2.
- [5] SABOL, J. – ZUBKOVÁ, I. 2006. Kultúra a multikultúrnosť. Bratislava: Euroformes, s.r.o. 2006. 54 s. ISBN 80-89266-01-0.
- [6] TESÁŘ, F. 2007. Etnické konflikty. 2007. Bratislava: Portál. 251 s. ISBN 80-73-67097-9.
- [7] VAŇO, B. 2002. Prognóza vývoja rómskeho obyvateľstva v SR do roku 2025. Bratislava: Inštitút informatiky a štatistiky. Edícia Akty. 2001. 15 s. 35-2002-A/10

Adresa autorov

Zuzana Poláková, Ing. PhD.
Tr. A. Hlinku 2
949 76 Nitra
Zuzana.Polakova@uniag.sk

Peter Obtulovič, doc., Ing., CSc.
Tr. A. Hlinku 2
949 76 Nitra
Peter.Obtulovic@uniag.sk

Oblúkový efekt korešpondenčnej analýzy Arch effect in Correspondence Analysis

Lucia Potisková

Abstract: Contingency table contains information about the joint distribution of two variables. For a large number of categories of these variables, the frequency matrix can be hard to interpret. Correspondence analysis is a tool for demonstrating relationships between row and column categories in lower dimension by an ordination map. In some cases there can appear a curious effect, when one can see a curve in a plot. It is called arch or horseshoe. After searching for a source of this, we discovered that it is caused by correlation between row and column categories.

Key words: correspondence analysis, chi-squared, arch effect, correlation

Kľúčové slová: korešpondenčná analýza, chí-kvadrát, oblúkový efekt, korelácia

JEL classification: C65

1. Úvod

Nech X a Y sú dve náhodné premenné s diskretným/kategorickým¹ rozdelením, $X \in \mathbf{X} = \{X_1, X_2, \dots, X_n\}$, $Y \in \mathbf{Y} = \{Y_1, Y_2, \dots, Y_p\}$. Nech náhodný výber (X, Y) má rozdelenie $p_{X,Y}(X, Y)$. Potom frekvencie výskytu i -tej kategórie ($i = 1, 2, \dots, n$) premennej X a j -tej kategórie ($j = 1, 2, \dots, p$) premennej Y , ozn. F_{ij} sú odhadom tohto združeného rozdelenia

$$F_{ij} = \hat{p}_{X,Y}(X = x_i, Y = y_j) \quad \text{pre } i = 1, 2, \dots, n \text{ a } j = 1, 2, \dots, p. \quad (1)$$

Frekvencie F_{ij} tvoria kontingenčnú tabuľku F o veľkosti $n \times p$. Korešpondenčná analýza sa zaoberá zobrazením vzťahov kategórií dvoch premenných, teda kategóriami náhodných premenných X, Y reprezentovanými riadkami a stĺpcami kontingenčnej tabuľky. Cieľom je zobrazenie tabuľky v 1D, 2D alebo 3D pri minimálnej strate informácií, keďže pri veľkom počte kategórií je veľmi ťažké interpretovať ich vzťahy. Preto je korešpondenčná analýza tak využívaná.

2. Princíp korešpondenčnej analýzy

Redukciou dimenzie je táto metóda veľmi podobná analýze hlavných komponentov (angl. Principal Component Analysis = PCA), avšak namiesto variancie rozkladá mieru asociácie (χ^2 štatistika). Za predpokladu, že by boli riadky a stĺpce nezávislé, mali by hodnoty pozorovaných frekvencií v bunke (i, j) tabuľky F nadobúdať hodnoty

$$E_{ij} = \frac{F_{i+} F_{+j}}{F_{++}} F_{++} = \frac{F_{i+} F_{+j}}{F_{++}} \quad (2)$$

kde $F_{i+} = \sum_{j=1}^p F_{ij}$, $F_{+j} = \sum_{i=1}^n F_{ij}$, $F_{++} = \sum_{i=1}^n \sum_{j=1}^p F_{ij} = \sum_{i=1}^n F_{i+} = \sum_{j=1}^p F_{+j}$.

Porovnaním nameraných hodnôt a tých očakávaných za predpokladu nezávislosti, dostávame štatistiku

$$t = \sum_{i=1}^n \sum_{j=1}^p \frac{(F_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i=1}^n \sum_{j=1}^p c_{ij}^2, \quad (3)$$

ktorá má za spomínaného predpokladu $\chi_{(n-1)(p-1)}^2$ rozdelenie a kde potom

¹môže ísť aj o premennú so spojitým rozdelením, avšak realizácie sú rozdelené do intervalov

$$c_{ij} = \frac{F_{ij} - E_{ij}}{\sqrt{E_{ij}}}. \quad (4)$$

Nech c_{ij} sú prvky matice C s hodnotou R . Na maticu C sa vieme pozrieť z pohľadu riadkov či stĺpcov ako na:

- n bodov v \mathbb{R}^p , kde súradnice bodu $c_i^T = [c_{i1}, \dots, c_{ip}] \in \mathbb{R}^p, i = 1, 2, \dots, n$ predstavujú i -ty riadok matice C
- p bodov v \mathbb{R}^n , kde súradnice bodu $c_j = [c_{j1}, \dots, c_{jn}] \in \mathbb{R}^n, j = 1, 2, \dots, p$ predstavujú j -ty stĺpec matice C .

Chceli by sme aproximovať riadkový a stĺpcový priestor nejakými menšími podpriestormi pri minimálnej strate informácií, ideálne podpriestormi \mathbb{R}, \mathbb{R}^2 alebo \mathbb{R}^3 , teda priamkou, plochou alebo priestorom. Kritériom výberu bude minimalizácia súčtu štvorcov odchýlok bodov voči ich projekciám na podpriestor s nižšou dimenziou. Riešenie tohto problému vedie k bázovým vektorom $\gamma_1, \gamma_2, \dots, \gamma_R$ a $\delta_1, \delta_2, \dots, \delta_R$ (hovoríme im faktorové osi) a k súradniciam projekcií $[C\delta_1, C\delta_2, \dots, C\delta_R]$, resp. $[C^T\gamma_1, C^T\gamma_2, \dots, C^T\gamma_R]$. Spomenuté vektory vieme získať aj pomocou SVD rozkladu (angl. Singular Value Decomposition)

$$C = \Gamma \Lambda \Delta^T, \quad (5)$$

kde matica Γ je zložená z vlastných vektorov $\gamma_1, \gamma_2, \dots, \gamma_R$ matice CC^T a Δ je zložená z vlastných vektorov $\delta_1, \delta_2, \dots, \delta_R$ matice C^TC . Matica Λ je diagonálna matica s prvkami $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_R}$, kde $\{\lambda_k\}_{k=1}^R$ sú zostupne zoradené vlastné čísla matíc CC^T a C^TC . Vďaka orthonormalite stĺpcov matice Γ a Δ a (5) vieme upravovať ďalej:

$$t = \sum_{i=1}^n \sum_{j=1}^p c_{ij}^2 = \text{tr}(CC^T) = \text{tr}(\Gamma \Lambda \Delta^T (\Gamma \Lambda \Delta^T)^T) = \text{tr}(\Lambda^2 I_R) = \sum_{k=1}^R \lambda_k$$

(6)

Tento vzťah nám ukazuje, že rozklad matice C pomocou singulárnych hodnôt rozkladá vlastne celkovú χ^2 hodnotu. Označme $\sum_{k=1}^R \lambda_k$ pojmom celková inercia. Pri aproximácii menšími podpriestormi dimenzie m berieme do úvahy iba príslušný počet vlastných čísel. Kvalitu takejto aproximácie nám popisuje percento vysvetlenej inercie

$$\tau_m = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_m}{\lambda_1 + \lambda_2 + \dots + \lambda_R}. \quad (7)$$

Snažíme sa teda o výber podpriestoru, v ktorom dosiahneme kompromis medzi veľkosťou dimenzie, v ktorej sme schopní interpretovať vzťahy medzi dvoma kategorickými premennými tabuľky F a maximalizáciou percenta vysvetlenej celkovej inercie τ_m .

Pozície riadkových a stĺpcových kategórií v redukovanom priestore sú vážené projekcie $C\delta_k$ resp. $C^T\gamma_k$:

$$r_k = A^{-1/2} C \delta_k = \sqrt{\lambda_k} A^{-1/2} \gamma_k \quad (8)$$

$$s_k = B^{-1/2} C^T \gamma_k = \sqrt{\lambda_k} B^{-1/2} \delta_k \quad (9)$$

kde $A_{(n \times n)} = \text{diag}(F_{i+})$ a $B_{(p \times p)} = \text{diag}(F_{+j})$ a druhý spôsob vyjadrenia v oboch vzťahoch je dôsledkom duality medzi podpriestormi riadkov a stĺpcov. Dôvodom vyváženia marginálnymi riadkovými frekvenciami v predchádzajúcich vzťahoch je, že získané projekcie sú teraz na každej z osí $k = 1, \dots, R$ centrované v nule. Ďalšími súradnicami sú

$$\hat{r}_k = \sqrt{\frac{F_{++}}{\lambda_k}} r_k = \sqrt{F_{++}} \mathbf{A}^{-1/2} \gamma_k \quad (10)$$

$$\hat{s}_k = \sqrt{\frac{F_{++}}{\lambda_k}} s_k = \sqrt{F_{++}} \mathbf{B}^{-1/2} \delta_k \quad (11)$$

a tie sú skonštruované tak, aby r_k boli centroidmi (váženými priermi) \hat{s}_k a \hat{r}_k boli centroidmi s_k . Podľa toho sú rozdelené aj jednotlivé typy škálovania, uvedieme dva najčastejšie využívané:

- Škálovanie typu 1 - vykresľujeme združený graf (angl. joint plot) pre riadkové profily r_k ako centroidy stĺpcových profilov \hat{s}_k . Takéto škálovanie teda používame v prípade, ak cieľom nášho záujmu sú predovšetkým riadkové kategórie.
 - Škálovanie typu 2 - vykresľujeme združený graf pre stĺpcové profily s_k ako centroidy riadkových profilov \hat{r}_k . Takéto škálovanie používame v prípade, ak cieľom nášho záujmu sú vzťahy medzi jednotlivými stĺpcovými kategóriami.
- Vzdialenosti medzi bodmi sú aproximáciami (ak je $\tau_m < 1$) ich χ^2 vzdialeností.

3. Interpretácia

Výsledkom tohto zobrazenia je korešpondenčná mapa. Tú interpretujeme v závislosti od typu škálovania. Ak sme zvolili škálovanie typu 1, tak sledujeme vzdialenosti medzi bodmi predstavujúce riadkové kategórie. Ak majú podobný profil výskytu v jednotlivých stĺpcových kategóriách, tak budú na korešpondenčnej mape vykreslené blízko seba. Ďalej, keďže stĺpcové kategórie majú koordináty také, že riadkové kategórie sú ich váženými priermi, tak ak má sledovaná riadková kategória vysoký počet výskytov v istej stĺpcovej kategórii, tak práve táto stĺpcová kategória bude zobrazená veľmi blízko sledovanej riadkovej kategórie. Analogicky to funguje v škálovaní typu 2, kde však svoju pozornosť upriamujeme na stĺpcové kategórie.

Príklad: sledujeme frekvenciu výskytu rôznych druhov motýľov v rôznych typoch prostredí. Naša kontingenčná tabuľka má tvar *prostredia* \times *druhy*. Pri škálovaní typu 1 sledujeme rôzne typy prostredia a vzťahy medzi nimi z pohľadu výskytu motýľov - napr. dalo by sa predpokladať, že keďže mnohé motýle sú veľmi citlivé na prostredie, oblasti s takmer rovnakým percentom zelene budú zobrazené blízko seba, naopak lúka a oblasť v blízkosti fabriky budú zobrazené na opačných stranách grafu. Pri škálovaní typu 2 sú cieľom nášho záujmu vzťahy medzi jednotlivými druhmi motýľov z pohľadu výskytu v rôznych prostrediach: ak nejaké druhy motýľov majú podobný profil prostredí, v ktorých sa vyskytujú, budú zrejme zobrazené na grafe veľmi blízko seba.

4. Oblúkový efekt

V autoritatívnej monografii [4], str. 465, je oblúkový efekt (angl. *arch effect*) pre ekologický príklad opísaný ako dôsledok unimodálneho rozdelenia kategórií pozdĺž environmentálnych gradientov². Pod environmentálnym gradientom si môžeme predstaviť napríklad zmenu nadmorskej výšky a pozdĺž gradientu môžeme sledovať napríklad výskyt rôznych druhov rastlín. Nech by bolo rozmiestnenie jednotlivých stanovíšť každých 50 výškových metrov, čiže by išlo o ich lineárne usporiadanie. Dá sa očakávať, že výskyt každej z rastlín bude mať unimodálne rozdelenie v závislosti od nadmorskej výšky, keďže každá rastlina má istú ekologickú niku (nároky na ekosystém výskytu), ktorú preferuje. Autori tvrdia, že takýto prípad sa prejaví na dvojrozmernej korešpondenčnej mape oblúkom.

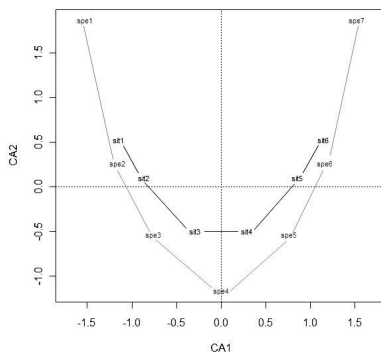
²Environmentálny gradient je postupná zmena neživých faktorov v priestore alebo v čase.

Očakávajú tiež, že lineárne usporiadanie by malo byť vidno aj na grafe.

Príklad: sledujeme výskyt 7 druhov (ozn. *species*) na 6 stanovištiach v rôznych nadmorských výškach (ozn. *sites*). Nech by bol výskyt taký, ako v tabuľke 1.

Tabuľka 1: Modelový príklad pre oblúkový efekt

	spe1	spe2	spe3	spe4	spe5	spe6	spe7
sit1	1	1	1	0	0	0	0
sit2	1	1	1	1	0	0	0
sit3	0	1	1	1	1	0	0
sit4	0	0	1	1	1	1	0
sit5	0	0	0	1	1	1	1
sit6	0	0	0	0	1	1	1



Obrázok 1: Modelový príklad v dvozmernej korešpondenčnej mape pri škálovaní 1

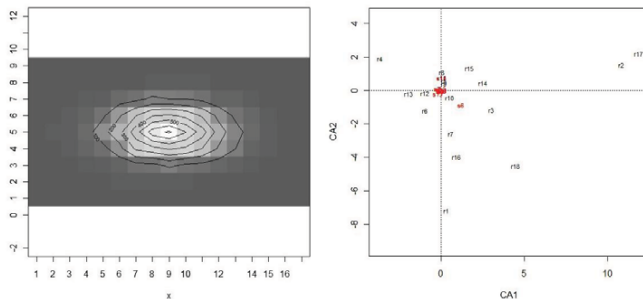
Pri škálovaní typu 1, teda pri zameraní sa na riadkové kategórie (stanovišťa), by všetky kategórie okrem *sit1* a *sit6* mali mať, vzhľadom na svoje početnosti výskytov, rovnakú χ^2 vzdialenosť od centroidu (priemerného riadkového profilu). Kategórie *sit1* a *sit6* ich budú mať väčšie. Ak by ich mali rovnaké ako ostatné riadkové kategórie a niektorý druh by sa vyskytoval na oboch stanovištiach, očakávali by sme, že body reprezentujúce jednotlivé stanovišťa by sa zobrazili do kruhu so stredom v centroide. Avšak keďže *sit1* a *sit6* nemajú žiadne spoločné druhy, budú od seba „odtláčané“, ale zároveň budú „príťahované“ stanovišťami s výskytom rovnakých druhov rastlín a vytvorí oblúk. To isté platí aj pre stĺpcové kategórie, keďže ich vážené priemery sú riadkové kategórie. Rovnaký efekt by sa objavil pri akomkoľvek type škálovania. Spomenutý lineárny trend je vidno v usporiadaní podľa 1. faktorovej osi.

5. Príčina oblúkového efektu v CA

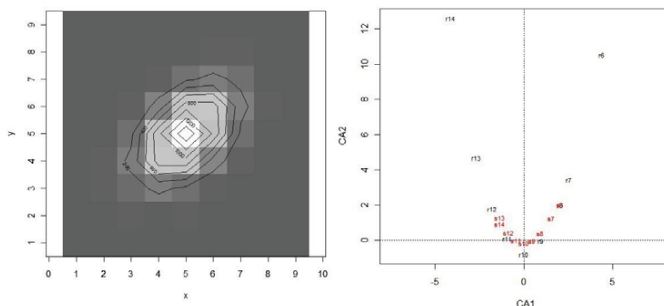
Domnievame sa, že vysvetlenie príčiny tohto efektu nie je správne, že efekt nie je zapríčinený unimodalitou, ale koreláciou medzi riadkovými a stĺpcovými profilmi. Zaujal nás tvar kontingenčnej tabuľky v modelovom príklade a snažili sme sa generovať rôzne kontingenčné tabuľky zo združeného normálneho rozdelenia s rôznymi parametrami. K získaniu kontingenčnej tabuľky sme museli vygenerované dáta rozdeliť do intervalov, rozkategorizovať. Uvedieme niekoľko príkladov na nasledovných obrázkoch. Jednotlivé grafy zobrazujú izolínie hustoty rozdelenia a korešpondenčnú mapu (škálovanie typu 2). Izolínie rozdelenia nám naznačujú tvar kontingenčnej tabuľky, keďže frekvencie v nej sú odhadom združenej pravdepodobnosti.

Takýmto spôsobom sme si vygenerovali niekoľko desiatok kontingenčných tabuliek a vykreslili sme si príslušné korešpondenčné mapy. Pri akejkoľvek kombinácii parametrov, ak

kovariancie boli nulové (Obrázok 2), oblúčkový efekt sa neobjavoval. Naopak, so zvyšujúcou sa kovarianciou sa body na korešpondenčnej mape formovali do jasnejšieho tvaru oblúku, podkovy (Obrázok 3). Pri nulových kovarianciách ide tiež o unimodálne rozdelenie, preto tento prípad môžeme považovať za kontrapríklad k predpokladu unimodality pre oblúčkový efekt.



Obrázok 2: Nulové kovariancie $\begin{pmatrix} X \\ Y \end{pmatrix} : N \left(\begin{pmatrix} 10 \\ 10 \end{pmatrix}, \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix} \right)$

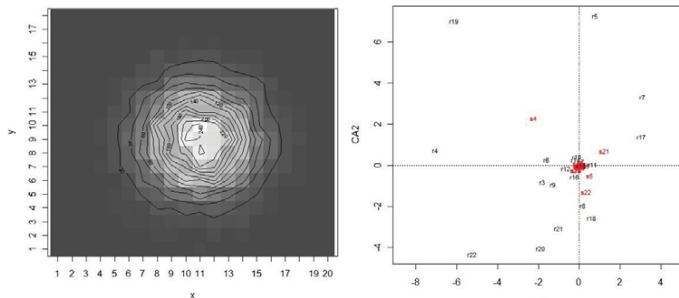


Obrázok 3: Nenulové kovariancie $\begin{pmatrix} X \\ Y \end{pmatrix} : N \left(\begin{pmatrix} 10 \\ 10 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right)$

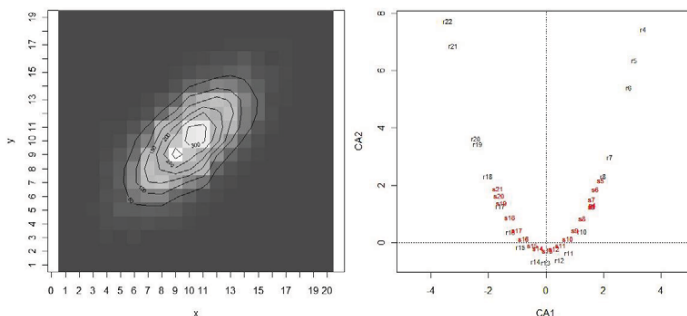
Predstavme si, že by sledovaná kategória nemala unimodálne, ale bimodálne rozdelenie. Na ukážku uvedieme jeden príklad bimodálneho rozdelenia s nulovými a nenulovými kovarianciami na obrázkoch 4 a 5. Aj v takomto prípade sa na korešpondenčnej mape ukázal oblúčkový efekt pri nenulových kovarianciách.

6. Záver

Simuláciami sa nám podarilo ukázať, že už desťročia staré tvrdenie, že oblúčkový efekt je spôsobený unimodalitou dát nie je pravdivé, ale že tento efekt je spôsobený závislosťou medzi riadkovými a stĺpcovými kategóriami. Niektorými odborníkmi bol vnímaný ako negatívny efekt a preto zaviedli detrendovanú korešpondenčnú analýzu, aby tento efekt odstránili, avšak pri vysvetlení príčiny oblúčkového efektu je teraz zrejmé, že takéto graf veľmi presne zachytáva reálnu situáciu a preto nie je potrebné detrendovanie.



Obrázok 4: Bimodálne rozdelenie, nulové kovariancie $\begin{pmatrix} X \\ Y \end{pmatrix} : N\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}\right) + N\left(\begin{pmatrix} 12 \\ 12 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}\right)$



Obrázok 5: Bimodálne rozdelenie, nenulové kovariancie $\begin{pmatrix} X \\ Y \end{pmatrix} : N\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 & 1.9 \\ 1.9 & 3 \end{pmatrix}\right) + N\left(\begin{pmatrix} 12 \\ 12 \end{pmatrix}, \begin{pmatrix} 3 & 1.9 \\ 1.9 & 3 \end{pmatrix}\right)$

7. Literatúra

- [1] GREENACRE, M. 2007. Correspondence Analysis in Practice. Chapman & Hall, 2007. 274 s. ISBN 1-58488-616-1.
- [2] HÄRDLE, W - HLÁVKA, Z. 2007. Multivariate Statistics: Exercises and Solutions. Springer Verlag, 2007. 368 s. ISBN 978-0-387-70784-6.
- [3] HÄRDLE, W. - SIMAR, L. 2003. Applied Multivariate Statistical Analysis. New York - Springer Verlag, 2003. 486 s. ISBN 3-540-03079-4.
- [4] LEGENDRE, P. - LEGENDRE, L. 1998. Numerical Ecology. Elsevier Science B.V., 1998. 853 s. ISBN 0-444-89249-4.

Adresa autora:

Lucia Potisková, Mgr.
 Fakulta prírodných vied Univerzity Mateja Bela
 Tajovského 40
 974 01 Banská Bystrica
 lucia.potiskova@gmail.com

Karcinóm endometria (CaE), analýza závislostí Carconoma endometrium (CaE), analysis of dependencies

Edward Radzo, Vanda Repiská, Ján Luha

Abstract: In this article we present Endometrial carcinoma originates in epithelial gland structures of the endometrium and it is the most common invasive neoplasia found in the in the developed countries. Numerous studies point to the distinct pathogenesis of various forms of this neoplasia. A dualistic model of endometrial tumorigenesis is currently recognized, broadly termed Type I and Type II. Endometrial carcinoma Type I occurs in the hyperplasia of endometrium. Type II usually occurs in atrophic endometrium. Mutation or deletion of the p53 tumorsuppressor gene on chromosome 17p13.3 is the most frequent gene mutation in human cancers. Polymorphism of gene TP53 can be harmful prognostic instrument of uterus neoplasias in future.

Key words: endometrial carcinoma, tumorsuppressor gene, endometrial adenocarcinoma, polymorphism

Kľúčové slová: endometriálny karcinóm, tumorsupresorový gén, endometriálny adenokarcinóm, polymorfizmus

JEL Classification: C1, C12, C14, I1.

1. Úvod

Prevalencia karcinómu endometria (CaE) v ženskej populácii má celosvetovo narastajúci trend. CaE je štvrtý najčastejší malígny nádor a zároveň, najbežnejšia gynekologická malignita u žien .

Podstatnú časť CaE tvorí endometrioidný histologický podtyp. Predpokladá sa, že endometrioidný podtyp CaE vzniká na podklade hyperplázie endometria (HE). Hyperplázia endometria sa môže vyskytovať v podobe jednoduchej HE bez atypie buniek až po komplexnú HE s prítomnou atypiou buniek. Niektorí autori zaraďujú medzi HE aj dobre diferencovaný karcinóm endometria. Endometrioidný histologický podtyp CaE sa vyskytuje u žien väčšinou v období perimenopauzy. Má nižší malígny potenciál, invázia do myometria je často minimálna a prienik do lymfatických priestorov je zriedkavejší. Vo všeobecnosti má dobrú prognózu, ako aj liečbu.

Vzhľadom na prognózu serózneho CaE je potrebné a užitočné identifikovať tento podtyp karcinómov aj v zmiešaných typoch CaE. DNA analýza génu *TP53* môže pomôcť identifikovať prítomnosť serózneho CaE, vzhľadom k tomu, že mutácie *TP53* sa vyskytujú hlavne v seróznom CaE a v karcinómoch endometria s vysokým stupňom histologickej diferenciácie (G III).

Histologický diagnostický nález skúmaných vzoriek bol od jednoduchej hyperplázie bez atypie buniek po endometrioidný CaE s výrazným prenikaním do myometria. Vzorky sme analyzovali priamym sekvenovaním, ako aj restričnou analýzou.

V súčasnosti sa DNA analýza nevyužíva v rutinej diagnostike, avšak prínos pre diferenciálnu diagnostiku, a tým aj liečbu CaE by mohol byť už v blízkej budúcnosti.

Výskumný súbor tvorilo 48 pacientiek, u ktorých bola potvrdená histopatologická diagnóza karcinómu endometria alebo prítomnosť prekancerózneho lézie. Pacientky sme rozdelili na dve vekové kategórie do 60 rokov (vrátane) a nad 60 rokov. V kategórii do 60 rokov bolo 20 pacientiek (41,7%) a v kategórii nad 60 rokov bolo 28 pacientiek (58,3%). V skúmaných vzorkách sa nachádzalo 87,5% vzoriek s CaE (42 prípadov) a 6 Prekancerózných vzoriek (12,5% prípadov). Rozdelenie podľa invazivity do myometria bolo: žiadna 33,3% (16 vzoriek), minimálna 54,2% (26 vzoriek) a vysoká 12,5% (6 vzoriek). 19 vzoriek (39,6% bolo heterozygotných, 26 (54,2% homozygotných a 3 (6,3%) bez poruchy génov.

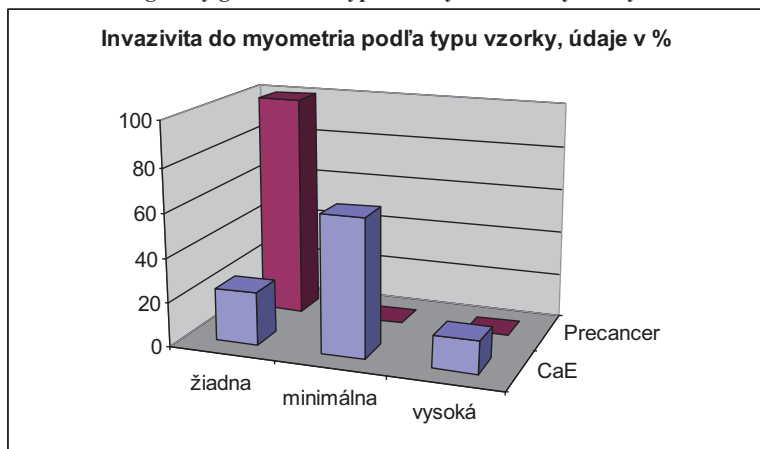
2. Analýza vzťahov podľa typu vzorky (CaE versus Precancer)

V príspevku sa venujeme analýze vzťahov typu vzorky (CaE, Precancer) s invazivitou do myometria, poruchov génov a vekom. Výsledky sú do určitej miery „poznačené“ malým počtom precancerózných vzoriek (iba 6 zo 48 skúmaných vzoriek).

Vzťah typu vzorky a invazivity do myometria prezentuje tabuľka 1. Uvedený vzťah plastickejšie prezentuje graf 1. Tento vzťah je štatisticky významný, P-hodnota Fisherovho exaktného testu je $P=0,002$. Precancerózne vzorky mali na 100% výskyt v skupine žiadna invazivita do myometria (všetkých 6 vzoriek). Vzorky s CaE mali v 23,8% žiadnu invazivitu do myometria (10 vzoriek), 61,9% minimálnu invazivitu do myometria (26 vzoriek) a 14,3% vysokú invazivitu do myometria (6 vzoriek).

Tabuľka 1: Kontingenčná tabuľka vzťahu typu vzorky a invazivity do myometria

			Invazia Invazivita do myometria			Total
			žiadna	minimálna	vysoká	
Typ	CaE	Count	10	26	6	42
		% within Typ	23,8%	61,9%	14,3%	100,0%
		% within Invazia Invazivita do myometria	62,5%	100,0%	100,0%	87,5%
		Adjusted Residual	-3,7	2,8	1,0	
Precancer		Count	6	0	0	6
		% within Typ	100,0%	,0%	,0%	100,0%
		% within Invazia Invazivita do myometria	37,5%	,0%	,0%	12,5%
		Adjusted Residual	3,7	-2,8	-1,0	
Total		Count	16	26	6	48
		% within Typ	33,3%	54,2%	12,5%	100,0%
		% within Invazia Invazivita do myometria	100,0%	100,0%	100,0%	100,0%

Graf 1: Kontingenčný graf vzťahu typu vzorky a invazivity do myometria

Typ vzorky a poruchu génov je v tabuľke 2. Tento vzťah nie je štatisticky signifikantný, P-hodnota Fisherovho exaktného testu je $P=0,127$, i keď faktické rozdiely v relatívnych početnostiach sú značné, malé početnosti však „spôsobiajú“, že vzťah nie je štatisticky signifikantný, ak by sme mali zachovanú štruktúru a počet vzoriek dvojnásobný, tak by uvedený vzťah bol vysoko štatisticky signifikantný. Pre hypoteticky zdvojnásobnený rozsah súboru by sme mali $P=0,006$.

Tabuľka 2: Kontingenčná tabuľka vzťahu typu vzorky a poruchy génov

		hetero			Total	
		heterozygot	homozygot	bez poruchy génov		
Typ	CaE	Count	14	25	3	42
		% within Typ	33,3%	59,5%	7,1%	100,0%
		% within hetero	73,7%	96,2%	100,0%	87,5%
		Adjusted Residual	-2,3	2,0	,7	
Precancer		Count	5	1	0	6
		% within Typ	83,3%	16,7%	,0%	100,0%
		% within hetero	26,3%	3,8%	,0%	12,5%
		Adjusted Residual	2,3	-2,0	-,7	
Total		Count	19	26	3	48
		% within Typ	39,6%	54,2%	6,3%	100,0%
		% within hetero	100,0%	100,0%	100,0%	100,0%

Podobne aj rozdiely podľa vekových kategórií nie sú štatisticky signifikantne odlišné, $P=1,000$ i keď „vidno“ „faktické“ rozdiely o 9,5%. Aby sme zistili aj štatisticky signifikantné rozdiely museli by sme rozsah súboru zväčšiť 22 krát.

Tabuľka 3: Kontingenčná tabuľka vzťahu typu vzorky a poruchy génov

			vek kat		Total
			do 60	nad 60	
Typ	CaE	Count	17	25	42
		% within Typ	40,5%	59,5%	100,0%
		% within vek_kat	85,0%	89,3%	87,5%
		Adjusted Residual	-,4	,4	
	Precancer	Count	3	3	6
		% within Typ	50,0%	50,0%	100,0%
		% within vek_kat	15,0%	10,7%	12,5%
		Adjusted Residual	,4	-,4	
Total		Count	20	28	48
		% within Typ	41,7%	58,3%	100,0%
		% within vek_kat	100,0%	100,0%	100,0%

3. Záver

Výsledky analýzy vzťahov typu vzorky ukázali vysokú štatistickú signifikantnosť s invazivitou do myometria, ďalej faktickú závislosť (i keď ešte nie štatisticky signifikantnú) závislosť s poruchou génov. Nezistili sme signifikantne odlišný vzťah s vekom pacientky. Mutácia alebo delécia v tumorsupresorovom géne TP53 na chromozóme 17p13.3 je najčastejšia genetická zmena v ľudských karcinómoch. Polymorfizmus v tumorsupresorovom géne TP53 môže byť užitočným prognostickým faktorom v rámci diferenciálnej diagnostiky karcinómov matrice.

4. Literatúra

- [1] Fleiss J.L., Levin B., Myunghee Ch.P. (2003): Statistical Methods for Rates and Proportions. 3ed. Wiley, New York 2003.
- [2] Greenberg R.S., Daniels S.R., Flanders W.D., Eley J.W., Boring J.R. III.: Medical Epidemiology. Fourth Edition, McGraw-Hill 2005.
- [3] Chajdiak J. (2007): Štatistika v exceli 2007. Statis, Bratislava 2009, ISBN 978-80-85659-49-8.
- [4] Kendall M. G., Stuart A. (1973): The Advanced Theory of Statistics, Volume 2, Inference and Relationship. Charles Griffin. London. Ruský preklad: Statističeskije vyvody i svjazi. Nauka Moskva 1973.
- [5] Lax SF. Molecular genetic pathways in various types of endometrial carcinoma: from a phenotypical to a molecular-based classification. Virchows Arch 2004; 444 (3): 213-223.
- [6] Liu FS. Molecular carcinogenesis of endometrial cancer. Taiwan J Obstet Gynecol 2007; 46 (1): 6-32.
- [7] Luha J. (1985): Testovanie štatistických hypotéz pri analýze súborov charakterizovaných kvalitatívnymi znakmi. Vydal Odbor Výskumu programov ČST a divákov v SR. Bratislava 1985.

- [8]Luha J.(2010a): Metodologické zásady záznamu dát z rozličných oblastí medicíny a zásady ich kontroly. FORUM STATISTICUM SLOVACUM 1/2010. SŠDS Bratislava 2010. ISSN 1336-7420.
- [9]Radzo E., Šišovský V., Sýkora P., Repiská V.: Molekulovo-genetická analýza tumorsupresorových génov PTEN a TP53 u pacientky s karcinómom endometria. In: Praktická gynekologie. - Roč. 13, č. 4 (2009), s. 225-227.
- [10] Radzo E., Sýkora P., Danihel L., Šišovský V., Repiská V.: DNA analýza a histologické podtypy karcinómu endometria = DNA analysis and histological subtypes of endometrial cancer. In: Praktický lékař. - Roč. 90, č. 3 (2010), s. 171-173.
- [11] Riffenburg R. H.: Statistics in Medicine, Second Edition. Academic Press 2005.
- [12] Rosner B.: Fundamentals of Biostatistics. Fifth Edition, Harvard University. Duxbury Press 2000.
- [13] Šišovský V., Danihel L., Novotný J., Bučeková B., Repiská V., Porubský J., Kopáni M., Černá A., Jakubovský J., Wawruch M-, Ševčovičová A., Manduch M. A., Helen Hatzibougias D., Kanibolotskiy A., Tsukamoto T.: Molecular genetics of endometrial carcinoma. In: Journal of Health Sciences Management and Public Health. - Vol. 9, No. 1 (2008), s. 111-118.
- [14] Tashiro H, Isacson C, Levine R, Kurman RJ, Cho KR, Hedrick L. p53 gene mutations are common in uterine serous carcinoma and occur early in their pathogenesis. Am J Pathol 1997; 150: 177-85.
- [15] Woodward M.: Epidemiology. Study design and data analysis. Second Edition. Chapman and Hall/CRC 2005.

Adresa autorov:

Edward Radzo, MUDr.

Ústav lekárskej biológie, genetiky a klinickej genetiky LF UK a UNB
Sasinkova 4, Bratislava
mudr.edwardradzo@gmail.com

Vanda Repiská, Doc., RNDr., CSc.

Ústav lekárskej biológie, genetiky a klinickej genetiky LF UK a UNB
Sasinkova 4, Bratislava
vanda.repiska@fmed.uniba.sk

Ján Luha, RNDr., CSc.

Ústav lekárskej biológie, genetiky a klinickej genetiky LF UK a UNB
Sasinkova 4, Bratislava
jan.luha@fmed.uniba.sk

Táto práca bola podporená Ministerstvom zdravotníctva SR v rámci projektu Imunohistochemická a molekulovo-genetická analýza karcinómu endometria 2007/28-UK-05 a grantom VEGA 1/0135/09.

Využitie štatistického softvéru C.H.I.C pri vyhodnotení didaktického experimentu

Apply of statistical software C.H. I.C in evaluation of didactical experiment

Lucia Rumanová, Gabriela Pavlovičová, Valéria Švecová

Abstract: V článku popisujeme štatistický softvér C.H.I.C. (Classification Hiérarchique Implicative et Cohésitive) a jeho použitie pri vyhodnotení didaktického experimentu. Prezentovaný experiment bol realizovaný na strednej škole v Palerme v spolupráci s talianskymi študentmi z Università degli Studi di Palermo. Spracovaniu jeho výsledkov predchádzala analýza a-priori v rámci TDS (teórie didaktických situácií).

Key words: test, students, goniometric functions, analyse a-priori, evaluation, similarity tree, implicative graph, implicative tree

JEL classification: C38, C88

1. Introduction

Analyse of particular didactical problem in the teaching processes is a research objective for one of the didactical school represented by Guy Brousseau or Yves Chevallard. Their research base of the theory of didactic situations is analysis of problem on particular levels of didactic situations (Brousseau, 1998; Chevallard, 1992). This theory formed our theoretical framework.

Also we want to present content analysis of students' test with using of statistical software C.H.I.C. (*Classification Hiérarchique Implicative et Cohésitive*)¹ that works with the frequencies of particular significant unites. This software represents the connection between quantitative and qualitative analysis, it makes possible to compare the similarity of didactical variables present in research, suggest the relations of coherence between variables and describes also the probability of realized implication between variables by probability rate of their realization. Apart from the relations between particular didactic variables this software allows also the comparison of relations between whole classes of didactic variables in three type of graph (similarity tree, implicative graph and implicative tree).

Our input information for analyze students' test were presence or absence of didactical variables of all strategies in students' work (value 1 for the presence or value 0 for the absence). The content analysis was realized for each problem from the test and we were tested 16 students of Liceo Socio – pedagogico Camillo Finocchiaro Aprile in Palermo.

2. Test

1. What is a positive angle?

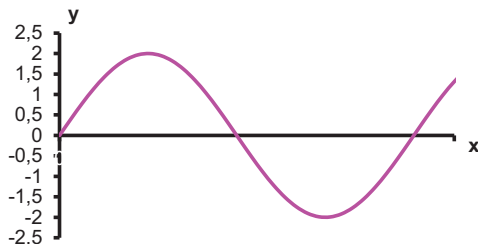
2. If $\alpha = 165^\circ$ then α in rad is:

- a) $\frac{\pi}{4}$ rad b) $\frac{\pi}{12}$ rad c) $\frac{11}{12}\pi$ rad d) $\frac{12}{11}\pi$ rad

3. Which one of four sets of angles represents the same angle? Choose the correct set and explain why it is right.

¹C.H.I.C (Version 3.1, 2003), Couturierat al. (2003)

- a) $90^\circ, 440^\circ, 810^\circ$ b) $45^\circ, 405^\circ, 765^\circ$ c) $30^\circ, 750^\circ, 1100^\circ$ d) $85^\circ, 455^\circ, 815^\circ$
4. Which is a maximum value of the function $y = \sin x$ and what is a correspondent angle?
5. Draw a part of graph of function $y = \sin x$ on interval $\left[-\frac{3}{2}\pi, \pi\right]$.
6. Choose the correct formula of the drowned graph.
- a) $y = 2 \cdot \sin x$ b) $y = 2 + \sin x$ c) $y = \sin 2x$ d) $y = \sin(x + 2\pi)$



Picture 1: The graph for problem 6

7. Draw the graph of function $y = \frac{1}{2} + \sin x$ (only one period).
8. Draw the graph of function $y = \sin\left(x - \frac{\pi}{2}\right)$ (only one period).
9. On the picture is drowned a graph of function $y = 3 \cdot \sin 2x$. Fill in the axes x, y and find the coordinates of signed point A.



Picture 2: The graph for problem 9

3. Analysis a – priori of the test

Before we gave the test to students we wrote analysis a – priori in which we tried to foresee the students' thinking and also their knowledge of the subject. These helped us to analyses the possible students' solutions. For example, to the first question we made out two points and we signed them A1, A2.

A1: He/she knows what is an angle.

A2: He/she doesn't know what is the orientation of the angle.

B1: He/she knows how to transform degrees into radians.

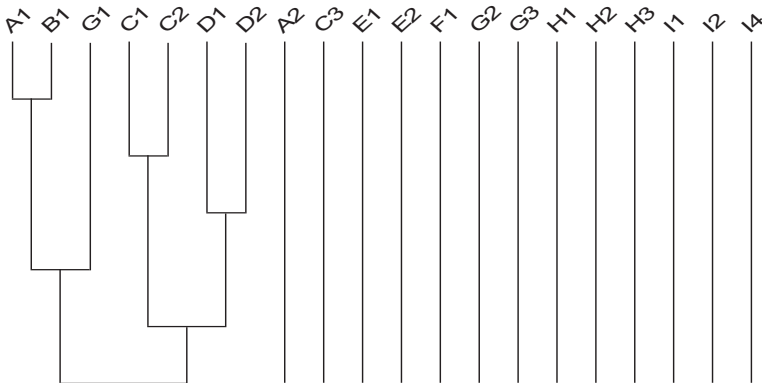
C1: He/she realizes that the period is 360° (2π).

C2: He/she knows to express the angle in form $\alpha = \alpha_0 + k \cdot 360^\circ$ ($\alpha = \alpha_0 + k \cdot 2\pi$).

- C3: He/she doesn't find the correct set of angles.
 D1: He/she knows the maximum (minimum) value of the function $y = \sin x$.
 D2: He/she determines the correspondent angle correctly.
 E1: He/she is able to draw the graph of the function $y = \sin x$.
 E2: He/she draws the graph of the function $y = \sin x$ on given interval.
 F1: He/she chooses the correct formula of the drowned graph.
 G1: He/she recognizes that the graph of the function should be moved in the direction of the axis y .
 G2: He/she realizes correctly the range of value.
 G3: He/she sketches the graph of given function rightly.
 H1: He/she recognizes that the graph of the function should be moved in the direction of the axis x .
 H2: He/she knows by what value should be the graph moved.
 H3: He/she sketches the graph of given function correctly.
 I1: He/she plots the axis x rightly.
 I2: He/she plots the axis y suitably.
 I3: He/she determines the period of the function correctly.
 I4: He/she realizes correctly the range of value.
 I5: He/she writes correct coordinates of signed point A.

4. Evaluation by statistical software CHIC

Similarity tree



Picture 3: The similarity tree by CHIC

Similarity tree represents the similarities between some arguments in analysis a-priori. By evaluation of experiment's results the most significant are the first two levels in the graph, the others are irrelevant.

From our similarity tree we can see:

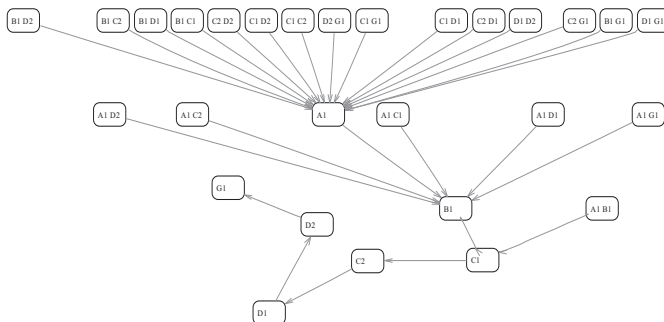
- The strongest similarity is between arguments $A1$ and $B1$ in analysis a-priori ($A1$: He/she knows what is an angle, $B1$: He/she knows how to transform degrees into radians).

- Not so strong similarities are between arguments $C1$ and $C2$ ($C1$: He/she realizes that the period is 360° (2π), $C2$: He/she knows to express the angle in form: $\alpha = \alpha_0 + k \cdot 360^\circ$).

($\alpha = \alpha_0 + k \cdot 2\pi$) and between $D1$ and $D2$ ($D1$: He/she knows the maximum (minimum) value of the function $y = \sin x$, $D2$: He/she determines the correspondent angle correctly).

- The second level of similarities is composed by following arguments/couples of arguments: $A1$, $B1$ and $G1$; $C1$, $C2$ and $D1$, $D2$.
- Between remaining arguments there isn't any similarity.

Implicative graph (Binomial distribution, number of variables 3)

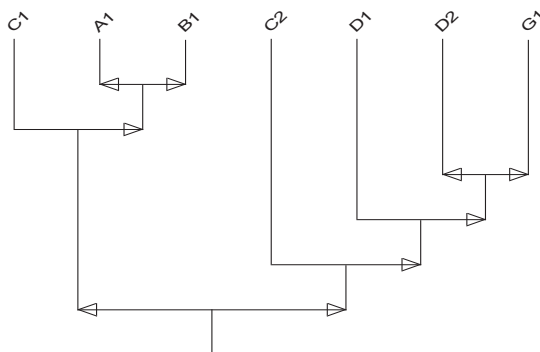


Picture 4: The implicative graph by CHIC

From this graph we can deduce that:

- The most couples of arguments implicate the argument $A1$ (for example: if the student has knowledge mentioned in arguments $B1$ (He/she knows how to transform degrees into radians) and $D2$ (He/she determines the correspondent angle correctly), then he knows what is an angle (= $A1$)).
- The argument $A1$ and the couples of arguments: $A1$ and $D2$, $A1$ and $C2$, $A1$ and $C1$, $A1$ and $D1$, $A1$ and $G1$ implicate the argument $B1$.
- The argument $B1$ and the couple of arguments $A1$ and $B1$ implicate the argument $C1$ (He/she realises that the period is 360° (2π)). This information is the most important for us because from it results which knowledge should the student have to understand the meaning of the term *period*.

Implicative tree



Picture 5: The implicative tree by CHIC

Implicative tree represents the implications or the equivalencies between some arguments in analysis a-priori. By evaluation of experiment's results the most significant are the first two levels in the graph, the others are irrelevant.

From this tree we can read following information:

- The most important is the first level where we can see these two equivalencies $A1 \Leftrightarrow B1, D2 \Leftrightarrow G1$.
- The second level is created only by implications $C1 \Rightarrow (A1 \Leftrightarrow B1)$ and $D1 \Rightarrow (D2 \Leftrightarrow G1)$.

3. Conclusion

New view on quantitative analysis results of didactical experiment us provides statistical program C.H.I.C. which slowly gets in awareness of didactics. The results of experiments are representing in three mentioned graphs in which we can read for example similarity, equivalence or percentual intensity between each variables or sets of variables which they were definite in analysis a-priori of given problem. This issue is still planning to give more detail.

4. Literature

- [1] BROUSSEAU, G. 1998. Théorie des situations didactiques. Grenoble, La Pensée sauvage, 1998.
- [2] CHEVALLARD, Y. 1992. Concepts fondamentaux de la didactique: perspectives apporées par une approche anthropologique. Recherches en Didactique des Mathématiques, Vol. 12/1. Grenoble, La Pensée sauvage, 1992.
- [3] KOHANOVÁ I. 2006. Didactical situation in specific conditions. In: Acta Didactica Universitatis Comenianae - Mathematics, Issue 6. Bratislava : Comenius University, 2006. ISBN 80-223-2138-9
- [4] TRENČANSKÝ, I. - RUMANOVÁ, L. - BEREKOVÁ, H. - HRÍBIKOVÁ, I. - REGECOVÁ, M. 2001. Slovník teórie didaktických situácií, 1. časť. In: Zborník príspevkov na seminári z teórie vyučovania matematiky, No. 4. Bratislava : Univerzita Komenského, 2001. ISBN 80-223-1704-7, p. 95-103.
- [5] TROVATO M. - MARCHIONI A. 1999. Moduli di matematica 2C. Minerva Italica, Milano.

Adresa autorov:

Gabriela Pavlovičová, PaedDr., PhD.
Tr. A. Hlinku 1
949 74 Nitra
gpavlovicova@ukf.sk

Lucia Rumanová, PaedDr., PhD.
Tr. A. Hlinku 1
949 74 Nitra
lrumanova@ukf.sk

Valéria Švecová, PaedDr., PhDr., PhD.
Tr. A. Hlinku 1
949 74 Nitra
vsvecova@ukf.sk

Štatistika - vstupná brána poznania Statistics - a gateway of knowledge

Beáta Stehlíková, Dagmar Markechová

Abstract: The aim of this article is discussing the teaching of mathematical statistics and the statistics in future periods, dominated by information technology.

Key words: statistics in the future, teaching of mathematical statistics, information technology

Kľúčové slová: budúcnosť štatistiky, vyučovanie matematickej štatistiky, informačné technológie

JEL classification: C10

1. Úvod

Matematická štatistika je v súčasnosti štandardným nástrojom, bez ktorého sa moderná veda, technika a ekonómia nezaobídu. Dáva odpoveď na otázku ako vyhodnotiť namerané alebo inak empiricky získané údaje a ako interpretovať získané výsledky. Štatistické postupy sa dnes bežne využívajú v praxi a vo všetkých oblastiach výskumu. Z týchto dôvodov sa matematická štatistika vyučuje na vysokých školách nielen na matematických odboroch štúdia, ale základy matematickej štatistiky sa vyučujú aj na nematematických odboroch štúdia, a to rovnako na prírodovedných ako aj na spoločenskovedných odboroch.

V predkladanom článku pojednávame o výučbe matematickej štatistiky a o budúcnosti štatistiky v období, v ktorom dominujú informačné technológie.

2. Budúcnosť štatistiky v období, v ktorom dominujú informačné technológie

Informácie sú v súčasnom období popri finančných, energetických a materiálových zdrojoch hlavným faktorom podmieňujúcim pokrok. Rozvoj informačných a komunikačných technológií má za následok hromadenie veľkého množstva údajov najrozličnejšieho charakteru. Štatistiku možno charakterizovať ako vedu, ktorej cieľom je vytážiť informácie a nové znalosti zo získaných údajov pri riešení problémov reálneho sveta.

Medzi najdôležitejšie úlohy, ktoré v súčasnom období stoja pred ľudstvom, patria tieto úlohy: katalogizovať a chrániť svetovú diverzitu druhov, odstrániť nedostatok vody a jej znečistenie, poskytovať čistú a dostupnú energiu, poskytovať všeobecné on-line vzdelávanie pre všetkých, eliminovať všetky hlavné infekčné a dedičné choroby, vykonávať medzinárodný dohľad nad kľúčovými technológiami. S rozvojom vedy a techniky nastupujú nové technológie, ktoré prinášajú nové typy experimentov a meraní, čo má za následok vznik nových typov údajov. Mení sa charakter údajov. Kým v minulosti údaje získané pri najrozličnejších experimentoch predstavovali malé až veľké súbory s veľkosťou $10^4 - 10^8$ bytov, v súčasnosti výsledky experimentov tvoria často obrovské až masívne súbory s veľkosťou $10^{10} - 10^{12}$ bytov. Na rozdiel od nízkej dimenzie súborov, ktorá bola charakteristická pre minulé obdobie, v súčasnosti prevláda vysoká dimenzia súborov. Mení sa aj charakter údajov. Zatiaľ čo v minulom období výsledky experimentov boli kvantitatívne, ordinálne resp. nominálne dáta, v súčasnosti popri numerických dátach sú často predmetom vyhodnotenia dáta nenumericke, napríklad obrazové. Tieto údaje obsahujú aj informácie o polohe a často sú neúplné. Tým vzniká pre štatistickú teóriu a jej metódy nová paradigma.

Informácie v dátach obsahujú spravidla istú neurčitost'. Existujú pritom rôzne zdroje neurčitosti a rôzne cesty ako neurčitost' vyjadriť. Vznikli viaceré teórie ako s neurčitost'ou pracovať. Jednou z takýchto teórií je teória fuzzy množín, ktorá umožňuje matematicky

popísať nepresné pojmy a pracovať s nimi. Jej rozvoj inicioval L. A. Zadeh, keď v článku [6] definoval pojem „fuzzy množina“ (fuzzy = hmlistý, neostrý, rozmazaný).

Mnohé skúmané javy existujú v priestore a sú viac či menej ovplyvňované svojím bezprostredným aj vzdialenejším okolím. Neexistujú izolovane od svojho okolia. Mnohé problémy majú teda priestorový charakter. Vznikajú nové typy dát, tzv. priestorové dáta. Nové typy dát vyžadujú nové metódy a nové techniky. Vzniká priestorová štatistika ako prostriedok na vyhodnocovanie priestorových dát.

Keďže realizácia štatistických výpočtov je časovo a numericky náročná, boli vyvinuté špeciálne štatistické softvéry, ktoré ich umožňujú ekonomizovať. V súčasnosti existuje veľké množstvo špeciálnych štatistických softvérov (napríklad STATISTICA, S, SPLUS, R, SAS), čo je hlavne pre aplikovanú štatistiku veľkým prínosom.

3. Vyučovanie matematickej štatistiky v období, v ktorom dominujú informačné technológie

Nové trendy a zmeny, ktoré priniesol súčasný rozvoj vedy, techniky a informačných technológií, je nevyhnutné premietnuť aj do výučby matematickej štatistiky na vysokých školách. Matematickú štatistiku je potrebné vyučovať v spojitosti s relevantnými aplikáciami s dôrazom na správnu interpretáciu výsledkov, pričom musí byť zachovaná matematická korektnosť. Zároveň je nutné, aby študenti pri riešení konkrétnych problémov zo štatistickej praxe využívali vhodný štatistický softvér.

Katedra matematiky Fakulty prírodných vied Univerzity Konštantína Filozofa v Nitre zabezpečuje výučbu matematickej štatistiky jednak pre študentov svojich študijných odborov, ale aj pre študentov ekológie, archeológie, sociálnej práce a masmediálnej komunikácie a tiež pre študentov doktorandského štúdia na Pedagogickej fakulte UKF v Nitre. Avšak aj študenti iných odborov používajú metódy matematickej štatistiky pri písaní svojich záverečných prác, či už bakalárskych, diplomových, rigorózných alebo dizertačných, v prípade že potrebujú štatisticky vyhodnotiť empirické údaje. Preto sú základy štatistiky v súčasnom období zaraďované do študijných programov aj ďalších študijných odborov, ktoré sa študujú na našej univerzite. Pre ilustráciu, v týchto dňoch prišla na našu katedru žiadosť o zabezpečenie výučby matematickej štatistiky pre študentov magisterského štúdia Pedagogiky na Pedagogickej fakulte UKF v Nitre.

Jedna z autoriek tohto článku pôsobila v predchádzajúcich rokoch na Fakulte ekonomiky a manažmentu Slovenskej Poľnohospodárskej Univerzity v Nitre. Podieľala sa na tvorbe učebných osnov ôsmich nových predmetov zameraných na matematickú štatistiku, ktoré boli následne zaradené do výučby študijných odborov FEM SPU v Nitre a v podstatnej miere sa podieľala aj na samotnom zabezpečení ich výučby. Sú to nasledujúce predmety:

1. Štatistická analýza priestorových údajov (bakalársky stupeň štúdia)
2. Priestorová štatistika (inžiniersky stupeň štúdia)
3. SAS ako nástroj kvantitatívnych analýz (inžiniersky stupeň štúdia)
4. Demografické modelovanie (inžiniersky stupeň štúdia)
5. Modelovanie regionálneho rozvoja (inžiniersky stupeň štúdia)
6. Finančná a poistná matematika (inžiniersky stupeň štúdia)
7. Modelovanie regionálneho rozvoja (doktorandské štúdium)
8. Štatistické metódy vo výskumnickej práci (doktorandské štúdium)

V rámci dištančného vzdelávania edukačného projektu BIOSTATEDUCA bol zaradený predmet Základný kurz zo štatistiky, ktorého cieľom bolo oboznámiť študentov so základmi popisnej štatistiky a vizualizáciou údajov.

Na doktorandskom stupni štúdia treba upozorniť študentov na špecifiká jednotlivých štatistických metód a oboznámiť ich s mnohorozmernými metódami. Pritom je potrebné držať krok so svetovým trendom v štatistike.

4. Záver

Rozvoj informačných a komunikačných technológií dal nové podnety pre rozvoj teórie matematickej štatistiky. Rigorózná matematická verifikácia je často nahrádzaná výpočtovými simuláciami. Rozvoj nových štatistických metód je a bude motivovaný novými problémami vychádzajúcimi z ľudského bádania. Budúcnosť štatistiky je preto v intenzívnejšej komunikácii štatistikov s výskumníkmi z iných odborov. Potreba štatistickej gramotnosti v dobe ovládanej vedou a technológiami sa stáva nevyhnutnosťou

5. Literatúra

- [1]ANDĚL, J. 2003. Statistické metody. Matfyzpress, vydavateľstvo MFF KU Praha, 2003, ISBN 80-85863-27-8.
- [2]MARKECHOVÁ, D. – TIRPÁKOVÁ, A. 2011. Štatistické metódy a ich aplikácie. FPV UKF v Nitre, 2011, ISBN 978-80-8094-807-8.
- [3]TIRPÁKOVÁ, A. – MARKECHOVÁ, D. 2008. Štatistika v praxi s popisom postupu práce v programe EXCEL. FPV UKF v Nitre, 2008, ISBN 978-80-8094-283-0.
- [4]TIRPÁKOVÁ, A. 2001. A few remarks about courses of applied statistics and about using computers for them. In: PRASTAN 2001 (Kočovce, 2001), s. 133-135.
- [5]TIRPÁKOVÁ, A. 2004. O problémoch vyučovania matematickej štatistiky na nematematických odboroch. In: Úloha učebnice vo vyučovaní matematiky (zborník príspevkov z vedeckého seminára), Nitra: FPV UKF, edícia Prírodovedec č. 157, (2004), s. 43-47, ISBN 80-8050-801-1.
- [6]ZADEH, L. A. 1965. Fuzzy Sets. Inform. and Control 8 (1965), s. 338-358.

Adresa autorov:

Beáta Stehlíková, prof., RNDr., CSc.
Fakulta ekonómie a podnikania PEVŠ
Tematínska 10
851 05 Bratislava
beata.stehlikova@uninova.sk

Dagmar Markechová, Doc., RNDr., CSc.
Katedra matematiky FPV UKF v Nitre
Trieda A. Hlinku 1
979 74 Nitra
dmarkechova@ukf.sk

Príspevok je financovaný z projektu FCVV FPV UKF Nitra.

Kvazinormy s obecnými podmínkami Quasi-norms with general constraints

Jakub Šácha

Abstract: This article is focused on theoretical and applicational problems with estimation of a discrete probability distributions from observed data. Theoretical aspects come out from minimizing Hellinger, Shannon and Pearson quasi-norm with given constraints. Compared to previous articles on this topic there are not only moment constraints, but also general linear constraints. At the end of the article there is some economic application.

Key words: estimation of probability distribution, quasi-norm, linear constraint.

Klíčové slova: odhad rozdělení pravděpodobnosti, lineární podmínka, kvazinorma.

JEL classification: C13

1. Úvod

Mezi nejdůležitější úlohy matematické statistiky patří nalezení rozdělení pravděpodobnosti pozorované náhodné veličiny či vektoru. Na základě pojmu f -divergence (vzdálenosti) dvou rozdělení je možné vyvodit postupy umožňující takové rozdělení odhadnout. Tyto postupy však musí obvykle respektovat další podmínky kladené na toto rozdělení. Jde nejčastěji o podmínky dané apriorním stanovením hodnot vybraných číselných charakteristik, např. střední hodnoty, rozptylu apod. Někdy však můžeme požadovat obecné lineární omezení. Naši základní ideou je najít takové rozdělení, které má nějaké požadované vlastnosti (splňuje zadané vedlejší podmínky) a je v jistém smyslu blízké vhodně zvolenému rozdělení. Přesněji jde o nalezení rozdělení, které je s takovým pevným rozdělením totožné při absenci vedlejších podmínek, ale s přidáváním podmínek se od tohoto pevného rozdělení postupně vzdaluje při současné minimalizaci zvolené f -divergence hledaného a daného pevného rozdělení.

2. Divergence a kvazinormy diskretních rozdělení pravděpodobnosti

Nechť funkce $f(u)$ je konvexní na $(0, \infty)$, striktně konvexní v $u=1$ a $f(1)=0$. f -divergencí rozdělení pravděpodobností \mathbf{p}, \mathbf{q} na diskretním pravděpodobnostním prostoru (Ω, Σ, P) rozumíme funkcionál $D_f(\mathbf{p}, \mathbf{q}) = \sum_x q(x) f\left(\frac{p(x)}{q(x)}\right)$, kde klademe $0f\left(\frac{0}{0}\right) = 0$, $0f\left(\frac{p}{0}\right) = pf(*)$ pro všechna $p \in (0, \infty)$ a $f(*) = \lim_{u \rightarrow \infty} \frac{f(u)}{u} \in \mathbb{R}^*$. Nechť $\mathbf{p} = (p_1, \dots, p_m)$ a $\mathbf{p}_0 = \left(\frac{1}{m}, \dots, \frac{1}{m}\right)$ pro $m > 1$ jsou diskretní rozdělení z pravděpodobnostního prostoru (Ω, Σ, P) , a D_f je f -divergence definovaná na daném prostoru. **Kvazinormou** rozdělení pravděpodobnosti $\mathbf{p} = (p_1, \dots, p_m)$ na (Ω, Σ, P) rozumíme f -divergenci $D_f(\mathbf{p}, \mathbf{p}_0)$. Pro odhady diskretních rozdělení za vedlejších momentových podmínek pomocí minimálních kvazinorem volíme:

a) Hellingerovu vzdálenost $D_{1/2}(\mathbf{p}, \mathbf{q})$, z níž získáme tzv. **Hellingerovu kvazinormu**

$$D(\mathbf{p}, \mathbf{p}_0) = \sum_{j=1}^m \left(\sqrt{p_j} - \sqrt{\frac{1}{m}} \right)^2 = 2 - \frac{2}{\sqrt{m}} \sum_{j=1}^m \sqrt{p_j},$$

b) I -divergenci $I(\mathbf{p}, \mathbf{q})$, z níž získáme tzv. **Shannonovu kvazinormu**

$$S(\mathbf{p}, \mathbf{p}_0) = \sum_{j=1}^m \left(p_j \ln p_j - \frac{1}{m} \ln \left(\frac{1}{m} \right) \right) = \sum_{j=1}^m p_j \ln p_j + \ln m,$$

c) χ^2 -divergenci $\chi^2(\mathbf{p}, \mathbf{q})$, z níž získáme tzv. **Pearsonovu kvazinormu**

$$P(\mathbf{p}, \mathbf{p}_0) = \frac{1}{m^2} \sum_{j=1}^m \frac{1}{p_j} - 1.$$

3. Odhady diskrétních rozdělení pravděpodobnosti

Předpokládáme, že pozorovaná diskrétní náhodná veličina X , jejíž rozdělení pravděpodobnosti $\mathbf{p} = (p_1, \dots, p_m)$ chceme odhadnout, nabývá nejvýše konečně mnoha navzájem různých reálných hodnot x_j^* s neznámými pravděpodobnostmi $p_j = P(X = x_j^*)$, $j = 1, \dots, m$, $m > 1$. Pozorováním náhodné veličiny X získáme statistický soubor (x_1, \dots, x_n) a jeho rozříděním dostaneme rozříděný statistický soubor $\left(\left(x_1^*, \frac{f_1}{n} \right), \dots, \left(x_m^*, \frac{f_m}{n} \right) \right)$, kde f_j je absolutní četnost pozorované hodnoty x_j^* . Dále předpokládáme, že $n > m$ a $f_j > 0$ pro všechna $j = 1, \dots, m$.

Pro odhad rozdělení \mathbf{p} požadujeme, aby toto rozdělení navíc splňovalo nějaké zadané podmínky, jejichž počet je $K - 1$. Mezi tyto podmínky zařazujeme zřejmou podmínku $\sum_{j=1}^m p_j = 1$. Hledáme pak takové rozdělení \mathbf{p} , které má minimální kvazinormu $D_f(\mathbf{p}, \mathbf{p}_0)$ za těchto dodatečných podmínek.

V předcházících článcích s touto tématikou jsme se zabývali momentovými podmínkami $\sum_{j=1}^m p_j x_j^{*k} = M_k$, $k = 0, \dots, K$. Jedná se o lineární podmínky v proměnných p_j s konkrétními koeficienty na levých stranách, a sice mocninami pozorovaných hodnot x_j^{*k} [4], [5], [6]. Tento příspěvek je zaměřen na obecné lineární podmínky

$$\sum_{j=1}^m p_j a_{kj} = b_k, \quad k = 0, \dots, K.$$

zapsáno maticově

$$\mathbf{A}\mathbf{p} = \mathbf{b},$$

$$\text{kde } \mathbf{A} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \dots & a_{Km} \end{pmatrix} \text{ je matice typu } K + 1 \text{ krát } m \text{ a } \mathbf{b} = \begin{pmatrix} 1 \\ b_1 \\ b_2 \\ \vdots \\ b_K \end{pmatrix} \text{ je sloupcový vektor}$$

pravých stran omezení.

Rozdělení pravděpodobnosti $\mathbf{p} = (p_1, \dots, p_m)$ pozorované diskrétní náhodné veličiny X má na pravděpodobnostním prostoru (Ω, Σ, P) , kde $\Omega = \{x_1^*, \dots, x_m^*\}$, $m > 1$, a Σ je množina všech podmnožin Ω , tzv. **minimální kvazinorma za K obecných lineárních podmínek $\mathbf{A}\mathbf{p} = \mathbf{b}$** , jestliže tato kvazinorma nabývá minimální hodnotu pro danou matici \mathbf{A} vektor \mathbf{b} . Jestliže $K < m - 1$, pak obdržíme pro minimální

$$a) \text{ Hellingerovu kvazinormu } p_j = \frac{1}{m \left(\sum_{k=0}^K \lambda_k x_j^{*k} \right)^2},$$

$$b) \text{ Shannonovu kvazinormu } p_j = \exp \left(-1 - \sum_{k=0}^K \lambda_k x_j^{*k} \right),$$

$$c) \text{ Pearsonovu kvazinormu } p_j = \frac{1}{m \left(\sqrt{\sum_{k=0}^K \lambda_k x_j^{*k}} \right)},$$

$j = 1, \dots, m$, kde λ_k , $k = 0, \dots, K$, jsou Lagrangeovy multiplikátory pro Lagrangeovu funkci

$$\Lambda(\mathbf{p}, \boldsymbol{\lambda}) = D_f(\mathbf{p}, \mathbf{p}_0) + \sum_{k=0}^K \lambda_k \left(\sum_{j=1}^m p_j a_{kj} - b_k \right), \quad \boldsymbol{\lambda} = (\lambda_0, \dots, \lambda_K).$$

Lagrangeovy multiplikátory λ_k je možno určit pomocí nelineární soustavy rovnic odpovídající nulovému gradientu Lagrangeovy funkce anebo přímo aplikovat některou metodu nelineární optimalizace pro určení jejího minima.

4. Příklad

Předpokládejme, že první podmínkou ($k = 1$) v zápise

$$\sum_{j=1}^m p_j a_{kj} = b_k,$$

chceme zaručit rovnost geometrických průměrů pro pozorování a odhad, tj.

$$\prod_{i=1}^m x_i^{*p_i} = \prod_{i=1}^m x_i^{*f_i/n}.$$

Tuto nelineární podmínku můžeme snadno linearizovat logaritmováním

$$\sum_{i=1}^m p_i \ln x_i^* = \sum_{i=1}^m \frac{f_i}{n} \ln x_i^* \equiv G_1.$$

Spolu se samozřejmou podmínkou řádu 0 můžeme tuto podmínku zapsat maticovým zápisem $\mathbf{A}\mathbf{p} = \mathbf{b}$ volbou

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \ln x_1^* & \ln x_2^* & \dots & \ln x_m^* \end{pmatrix}$$

a

$$\mathbf{b} = \begin{pmatrix} 1 \\ G_1 \end{pmatrix}.$$

Pokud bychom chtěli zaručit ještě navíc stejnou variabilitu pro pozorování a odhad přidáme ještě další podmínku

$$\sum_{i=1}^m p_i \ln^2 x_i^* = \sum_{i=1}^m \frac{f_i}{n} \ln^2 x_i^* \equiv G_2$$

a v maticovém zápisu $\mathbf{A}\mathbf{p} = \mathbf{b}$ volíme

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \ln x_1^* & \ln x_2^* & \dots & \ln x_m^* \\ \ln^2 x_1^* & \ln^2 x_2^* & \dots & \ln^2 x_m^* \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ G_1 \\ G_2 \end{pmatrix}.$$

Jedná se rovnost "geometrických" momentů druhého řádu.

Mějme nyní konkrétní datový soubor o rozsahu $n = 100$, jehož rozříděním dostaneme následující tabulku.

x_i^*	0,8	1	1,2	1,4	1,6
f_i	17	35	27	14	7

Předpokládejme, že tato data vyjadřují koeficienty růstu nějaké veličiny za 100 časových období, např. měsíců. Je tedy věcně správnější pracovat s geometrickým průměrem (příp. obecně s geometrickými momenty vyšších řádů) než aritmetickým průměrem (příp. obecně s aritmetickými momenty vyšších řádů). Vypočítejme a srovnajme tedy odhady pomocí minimalizace Pearsonovy kvazinormy v prvním případě při použití klasických aritmetických momentových podmínek, jejichž použití není pro takováto data věcně správné a v druhém případě při použití správnějších geometrických momentových podmínek. V obou případech volíme rovnost momentů až do řádu $K = 1$ příp. $K = 2$.

Počet tříd je $m = 5$ rozsah $n = 100$, takže

$$M_1 = \frac{1}{100} \sum_{i=1}^m f_i x_i^* = 1,301,$$

$$M_2 = \frac{1}{100} \sum_{i=1}^m f_i x_i^{*2} = 1,096,$$

$$G_1 = \frac{1}{100} \sum_{i=1}^m f_i \ln x_i^* = 0,0913,$$

$$G_2 = \frac{1}{100} \sum_{i=1}^m f_i \ln^2 x_i^* = 0,0488$$

Pomocí optimalizačního nástroje Řešitel z Excelu pro určení minima Pearsonovy kvazinormy jsme získali výsledky v následujících tabulkách. První tabulka odpovídá omezením ve tvaru momentových podmínek pro $K = 1$, příp. $K = 2$ a druhá odpovídá omezením ve tvaru "geometrických" momentových podmínek, popsanych výše.

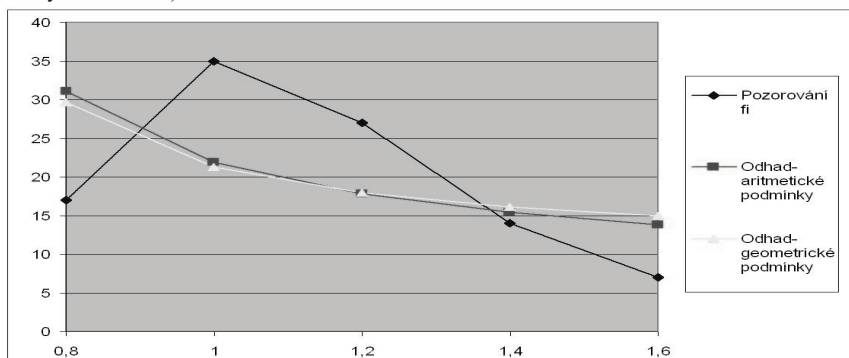
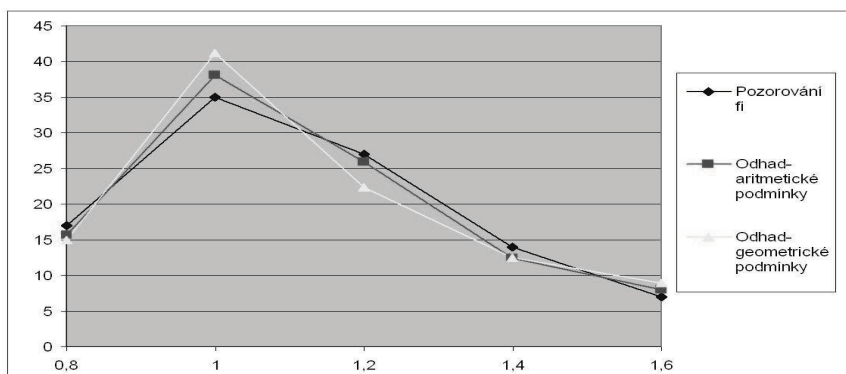
Tabulka 1: Výsledky pro aritmetické momentové podmínky

x_i^*	0,8	1	1,2	1,4	1,6
f_i	17	35	27	14	7
Odhad $K = 1$	31,1	21,9	17,8	15,4	13,8
Odhad $K = 2$	15,6	38,1	25,9	12,4	8,0

Tabulka 2: Výsledky pro geometrické momentové podmínky

x_i^*	0,8	1	1,2	1,4	1,6
f_i	17	35	27	14	7
Odhad 1	29,6	21,3	18,0	16,2	15,0
Odhad 2	15,0	41,2	22,4	12,5	8,9

Pro přehled uvedme ještě grafické znázornění výsledků. V prvním grafu je srovnáno pozorování s odhady při použití aritmetických a geometrických momentových podmínek pro $K = 1$ (rovnost průměrů). V druhém grafu je srovnání téhož pro $K = 2$ (rovnost průměrů a druhých momentů).

**Obrázek 1: Znázornění výsledků pro $K = 1$** **Obrázek 2: Znázornění výsledků pro $K = 2$**

5. Závěr

Empirický přístup k odhadům rozdělení pravděpodobnosti vyžaduje při řešení konkrétních úloh dostatečnou dávku zkušeností a nelze přitom spoléhat na profesionální statistické softwarové produkty, které navíc obsahují pouze nevelké množství různých typů rozdělení. V tomto článku je ukázáno zobecnění momentových podmínek na obecné lineární podmínky, které mají širší možnost použití. Jak je ukázáno v příkladě, hodí se např. pro data, u kterých je smysluplnější pracovat s geometrickým průměrem oproti aritmetickému.

6. Literatura

- [1]VAJDA, I.: Teória informácie a štatistického rozhodovania. Bratislava: Alfa, 1982.
- [2]KARPÍŠEK, Z., JURÁK, P.: Modelling of Probability Distribution with Maximum Entropy. In: MENDEL '01.7th International Conference on Soft Computing. Brno, 2001, pp. 232-239, ISBN 80-214-1894-X.
- [3]KARPÍŠEK, Z., SADOVSKÝ, Z., ŠÁCHA, J.: Pitman – Hellinger Test of Fit. In: 4th International Conference APLIMAT 2005 (part II). Bratislava, 2005, pp. 471- 478, ISBN 80-969264-2-X.
- [4]ŠÁCHA, J.; KARPÍŠEK, Z. Kvazinormy pro odhady diskretních rozdělení pravděpodobnosti. In *Moderní matematické metody v inženýrství - 3mi 2007*. Ostrava: JČMF, KMDG VŠB-TU, 2007. s. 276-280. ISBN: 978-80-248-1649-4.
- [5]ŠÁCHA, J.; KARPÍŠEK, Z.; JURÁK, P. Qusi-Norms for Discrete Probability Distribution Estimation. In *Risk, quality and reliability*. Ostrava, Czech republic: VŠB - Technical University of Ostrava, 2007. s. 179-185. ISBN: 978-80-248-1575-6.
- [6]KARPÍŠEK, Z.; JURÁK, P.; ŠÁCHA, J. Kvazinormy pro odhady diskretních rozdělení pravděpodobnosti. In *Sborník přednášek 8. národní konference Statistické dny v Brně*. Brno: Česká statistická společnost, 2006. s. 49-61. ISBN: 80-214-3214-4.

Adresa autora:

Jakub Šácha, Ing.
Ústav statistiky a operačního výzkumu
PEF MENDELU v Brně
Zemědělská 1
61300 Brno
jakub.sacha@post.cz

Analýza výkonov Puncového úradu Slovenskej republiky v rokoch 1970 – 2010 použitím štatistických metód Performances of the Assay Office of the Slovak republic in the years 1970 - 2010

Ján Tirpák

Abstract: Selection performance of the Assay Office was subdual statistical analysis namely application methods correlation time series and methods clustering analyses.

Key words: Assay control, jewellery, time series, cluster analysis

Kľúčové slová: puncová kontrola, šperky, časové rady, zhuková analýza

1. Úvod

Výkony Puncového úradu, ďalej PÚ podľa zákona spočívajú v zisťovaní a overovaní rýdzosti výrobkov a vecí z drahých kovov a to najmä zlata, striebra a platiny, pričom ich výsledkom sú úradné označenia tovaru puncovou značkou zodpovedajúcej zákonnej rýdzosti. Klenotnícke zliatiny používané na výrobu tovarov z drahých kovov musia mať predpísané zákonné rýdzosti a popri základnom drahom kove môžu obsahovať len kovy, ktoré sú v súlade s podmienkami ustanovenými zákonom o puncovníctve a skúšaní drahých kovov č. 10/2004 Z. z. V prípade tovaru, ktorý je vyrobený z registrovanej klenotníckej zliatiny sa skúška rýdzosti vykonáva nedeštruktívnym spôsobom na skúšobnom kameni alebo porovnaním so štandardom pomocou X – testu. Chemická skúška rýdzosti sa vykoná vždy, ak je tuzemský alebo cudzí tovar vyrobený z neregistrovanej zliatiny.

Puncový úrad SR na svojej internetovej stránke zverejnil štatistické údaje o svojej činnosti v niektorých prípadoch až za obdobie 1970 – 2010 [4]. Vybrané dáta sme podrobili štatistickej analýze a to aplikáciou metódy korelácie časových radov a metódy zhukovej analýzy.

2. Materiál a metódy

Vývoj výkonov PÚ zo štatistického hľadiska charakterizuje časový rad. Ak sledujeme viaceré časové rady, zaujíma nás, či medzi týmito časovými radmi neexistujú také súvislosti, ktoré by dovoľovali vysvetliť zmeny v jednom časovom rade zmenami v druhom časovom rade. Tieto súvislosti môžeme popísať pomocou koeficientov korelácie časových radov.

Pri skúmaní vzťahov medzi časovými radmi predpokladáme, že každý rad môžeme vyjadriť ako súčet pravidelnej a nepravidelnej zložky. Ak chceme zistiť, či medzi časovými radmi existuje určitý príčinný vzťah, nestačí skúmať iba celkovú vývojovú tendenciu, resp. sezónne kolísanie, pretože tieto faktory môžu mať zhodný priebeh, čo by skreslilo výsledný koeficient korelácie. Preto je potrebné skúmať, či neexistuje vzťah medzi náhodnými zložkami časových radov, na základe čoho by sme mohli predpokladať, že existuje skutočná príčinná závislosť medzi sledovanými časovými radmi [3].

Jednou z možností ako merať tesnosť závislosti časového radu je metóda diferencií [2]. Princípom tejto metódy je výpočet diferencií dostatočne vysokého radu, čím vylúčime ich trendovú zložku a ďalej teda pracujeme len s vypočítanými diferenciami, ktoré predstavujú určitým spôsobom pozmenenú reziduálnu zložku. V najjednoduchšom prípade, keď je možné popísať vývoj časový rad x_t lineárnou trendovou funkciou, je možné vylúčiť vplyv trendovej zložky pomocou prvých absolútnych diferencií $\Delta x_t = x_t - x_{t-1}$.

Koeficienty korelácie medzi diferenciami analyzovaných časových radov charakterizujú miery tesnosti medzi reziduálnymi zložkami týchto radov. Koeficient korelácie medzi diferenciami dvoch časových radov vypočítame podľa vzťahu:

$$r_{\Delta x_i \Delta y_i} = \frac{\sum \Delta x_i \Delta y_i - \frac{\sum \Delta x_i}{n} \cdot \frac{\sum \Delta y_i}{n}}{\sqrt{\left\{ \frac{\sum \Delta^2 x_i}{n} - \left(\frac{\sum x_i}{n} \right)^2 \right\} \left\{ \frac{\sum \Delta^2 y_i}{n} - \left(\frac{\sum y_i}{n} \right)^2 \right\}}}$$

Korelačný koeficient nadobúda hodnoty z intervalu $\langle -1, 1 \rangle$. Jeho hodnoty interpretujeme nasledovne: ak $r \doteq 1$, potom medzi znakmi X a Y existuje kladná lineárna závislosť, t.j. veľkým hodnotám znaku X zodpovedajú veľké hodnoty znaku Y a naopak. Ak $r \doteq -1$, potom medzi znakmi X a Y existuje záporná korelácia (výrazne protikladný vzťah). Veľkým hodnotám znaku X zodpovedajú malé hodnoty znaku Y a naopak. V prípade lineárnej nezávislosti je hodnota koeficienta korelácie $r = 0$. Hodnoty znakov X a Y sú v tomto prípade rozptýlené nezávisle od seba. Hodnota koeficienta korelácie môže byť rovná 0 aj v prípade, že medzi znakmi X a Y je iná štatistická súvislosť ako lineárna.

Tabuľka 1: Výkony PÚ SR vyjadrené v počtosti a hmotnosti opuncovaných zlatých a strieborných šperkov na Slovensku v rokoch 1970 – 1989

Rok	Au (ks)	Au (g)	Ag (ks)	Ag (g)
1970	188805	879739	125801	695684
1971	219557	955054	89708	588471
1972	232819	1014335	76772	501175
1973	249539	1071164	80943	513810
1974	194183	766895	132746	894360
1975	181913	889578	137519	952204
1976	244590	854252	152913	1067129
1977	262982	929987	175535	1258869
1978	267345	946224	320477	1387809
1979	276959	931947	235790	1931211
1980	165546	463593	286798	1169018
1981	227613	608721	282372	741640
1982	231758	624848	202571	636523
1983	234565	590019	245648	820671
1984	212464	683450	221290	670932
1985	236390	709912	218306	775697
1986	218668	676634	216882	643461
1987	200085	644770	185337	463361
1988	215402	684997	193887	593595
1989	218884	745002	196295	549107

Zdroj: [4]

Pre popísanie podobnosti časových radov použijeme aj metódu zhlukovej analýzy [2]. Metrika, ktorá vhodne popisuje podobnosť časových radov. Je daná vzťahom,

$d(x, y) = \sqrt{2(1-r(x, y))}$, kde $r(x, y)$ je korelačný koeficient medzi hodnotami časových radov x a y .

V príspevku analyzujeme údaje o výkonoch PÚ, konkrétne hmotností šperkov z platiny, zlata a striebra v rokoch 1970 - 1989, 1989 - 2010 a 1994 - 2010 a tiež údaje o skúškach rýdzosti drahých kovov v rokoch 1994 - 2010 [4]. V ďalšej analýze sme pracovali s logaritmi hodnôt.

3. Výsledky a diskusia

Analýzovali sme výkony PÚ, konkrétne mieru tesnosti väzby medzi hmotnosťami opuncovaných zlatých a strieborných šperkov PÚ za roky 1970-1989 (tab.1) a za roky 1980 - 2010 (tab.2).

Tabuľka 2: Výkony PÚ SR vyjadrené v početnosti a hmotnosti opuncovaných zlatých a strieborných šperkov na Slovensku v rokoch 1990 – 2010

Rok	Au (ks)	Au (g)	Ag (ks)	Ag (g)
1990	316054	1023987	161091	678233
1991	351841	933864	192721	1155317
1992	589122	1211601	76508	297431
1993	733134	2067597	75026	564922
1994	745997	1738305	63977	590662
1995	962773	2284103	67260	482911
1996	1064803	2860395	103046	442550
1997	1230601	2570780	100497	1044571
1998	1248082	3051659	251541	1766199
1999	993712	2305558	441403	2394695
2000	885083	1982442	517894	2495496
2001	919499	2203413	784408	3773011
2002	855398	2076365	970390	4719684
2003	779207	1913430	915902	5087141
2004	646376	1669325	511669	4135023
2005	621381	1681083	568618	5124233
2006	688446	1899980	754817	6960748
2007	859414	2327043	915685	7447134
2008	734918	2132392	708423	6119482
2009	548585	1604953	611196	4881229
2010	396148	1068304	431819	3196814

Zdroj: [4]

Podľa vzťahu (3) sme vypočítali hodnotu koeficienta korelácie medzi časovými radmi hmotností opuncovaných zlatých a strieborných šperkov ($r_{\Delta x, \Delta y} = 0,02302$) v rokoch 1970 - 1989. Na základe vypočítanej hodnoty koeficienta korelácie môžeme povedať, že medzi časovými radmi hmotností opuncovaných zlatých a strieborných šperkov v rokoch 1970 - 1989 nie je štatisticky významná korelačná závislosť.

Analogicky sme vypočítali tesnosť väzby ($r_{\Delta x, \Delta y} = 0,058145$) medzi uvedenými znakmi aj v období 1990 - 2010. Aj v tomto prípade medzi časovými radmi hmotností opuncovaných

zlatých a strieborných šperkov v rokoch 1990 - 2010 nie je štatisticky významná korelačná závislosť.

Výsledné hodnoty koeficientov korelácie poukazujú na skutočnosť, že v rokoch 1970 - 1989 zmeny hmotností opuncovaných šperkov zo zlata nevyvolávajú zmeny hmotností opuncovaných šperkov zo striebra a naopak, t.j. menia sa oba na seba nezávisle.

Ďalej sme pomocou koeficientov korelácie časových radov zisťovali mieru tesnosti väzby medzi hmotnosťami opuncovaných platinových, zlatých a strieborných šperkov PÚ za roky 1994 - 2010 (tab.3).

Tabuľka 3: Opuncované platinové, zlaté a strieborné šperky vyjadrené v hmotnosti za obdobie 1994 až 2010 v SR

Rok	Platina (kg)	Zlato (kg)	Striebro (kg)
1994	0,029	1738	591
1995	0,003	2284	483
1996	0,021	2860	443
1997	0,011	2571	1045
1998	0,310	3052	1766
1999	1,505	2306	2395
2000	0,548	1982	2495
2001	0,537	2203	3773
2002	0,623	2076	4720
2003	0,161	1913	5087
2004	0,411	1669	4135
2005	0,158	1681	5124
2006	0,262	1900	6961
2007	0,244	2327	7447
2008	0,245	2132	6119
2009	0,141	1605	5881
2010	0,567	1068	3197

Zdroj: [4]

Hodnoty koeficientov korelácie medzi časovými radmi hmotností opuncovaných platinových, zlatých a strieborných šperkov v rokoch 1994 - 2010 sú uvedené v nasledujúcej korelačnej matici (tab.4).

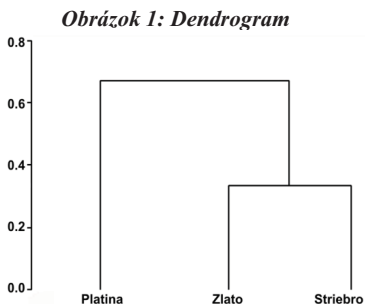
Tabuľka 4: Koeficienty korelácie pre prvé absolútne diferencie logaritmov hmotností opuncovaných platinových, zlatých a strieborných šperkov v SR

	Pt	Au	Ag
Pt	1		
Au	-0,03265	1	
Ag	0,063003	0,255004	1

Zdroj: vlastné vyhotovenie

Na základe vypočítaných hodnôt koeficientov korelácie (tab.4) môžeme povedať, že medzi časovými radmi hmotností opuncovaných zlatých a strieborných šperkov, platinových a zlatých šperkov a ani medzi časovými radmi hmotností opuncovaných strieborných a platinových šperkov v rokoch 1994 - 2010 nie je štatisticky významná korelačná závislosť.

Výsledné hodnoty koeficientov korelácie poukazujú na skutočnosť, že v rokoch 1994 - 2010 zmeny hmotností opuncovaných šperkov zo zlata nevyvolávajú zmeny šperkov zo opuncovaných striebra (a naopak, koeficienty korelácie sú symetrické), t.j. menia sa oba na seba nezávisle. Použitím zhlukovej analýzy sme dostali nasledovný dendrogram (obr.1).



Zdroj: vlastné vyhotovenie

Z obrázku vidíme, že podobnosť medzi časovými radmi je veľmi nízka, čo potvrdilo aj výsledky, získané pomocou metriky kvantifikujúcej podobnosť medzi časovými radmi.

4. Záver

Použitím štatistických metód sa zistilo, že časový priebeh hmotností opuncovaných zlatých šperkov v rokoch 1970 – 2010 nekoreluje s časovým priebehom hmotností opuncovaných strieborných šperkov. Teda výsledné hodnoty koeficientov korelácie poukazujú na skutočnosť, že v rokoch 1970 – 2010 zmeny hmotností opuncovaných šperkov zo zlata nevyvolávajú zmeny hmotností opuncovaných šperkov zo striebra a naopak, t.j. menia sa oba na seba nezávisle. K rovnakým záverom sme dospeli, keď sme sledovali súvislosti medzi časovými priebehmi hmotností opuncovaných šperkov z troch vzácnych kovov Pt, Au a Ag v rokoch 1994 – 2010. Aj tu sa potvrdilo, že podobnosť medzi časovými radmi hmotností šperkov z troch vzácnych kovov je veľmi nízka.

5. Literatúra

- [1]IVOR,J. - STEHLÍKOVÁ,B. - TIRPÁKOVÁ,A. – MARKECHOVÁ,D.: Je európsky trh s kokaínom jednotný? Forum Statisticum Slovacum 1/2009, s. 16-24 ISSN 1336-7420
- [2]MAHARAJ, E. A.: Clusters of time series. Journal of Classification, 17, 2000, 297- 314
- [3]SEGER, J.: Statistické metódy pro ekonomy průmyslu. SNTL, Praha 1988, 512 s.
- [4] <http://www.puncovyurad.sk/sk/Statistika.alej>

Adresa autora:

Ján Tirpák, RNDr., CSc.
 Katedra zoológie a antropológie FPV UKF
 Tr. A. Hlinku 1
 949 01 Nitra
jtirpak@ukf.sk

Výskyt sexuálnych porúch v SR – päťročná analýza pomocou dolovania dát Occurrence of sexual disorders in Slovakia - a five-year analysis with Data Mining

Andrej Trnka, Mária Kovářová, Ivan Dóci

Abstract: The authors address occurrence of sexual disorders in the population of the Slovak Republic. They worked with data issued by the National Health Information Centre of the SR for the period 5 years (2004-2008). Data mining was used in analysing the data. The results will serve the professional management of health care to improve it in the most exposed regions and in the risk groups.

Key words: sexual dysfunctions, age, gender, number of first examinations, outpatient care, data mining.

Kľúčové slová: sexuálne dysfunkcie, vek, pohlavie, počet prvých vyšetrení, ambulánna starostlivosť, dolovanie dát.

JEL classification: I10, C10

1. Úvod do problematiky

Podľa Medzinárodnej klasifikácie duševných porúch, 10 revízia, ktorá je platná od 1. januára 1994, diagnóza s kódom F52 má názov sexuálne dysfunkcie nevyvolané organickou poruchou alebo chorobou. Pojem sexuálna dysfunkcia zahŕňa rôzne stavy, kedy sa osoba nemôže podieľať na sexuálnom styku tak, ako si želala. Môže chýbať záujem o styk, uspokojenie zo styku alebo chýbajú fyziologické reakcie potrebné na úspešnú realizáciu styku (napríklad erekcia), alebo človek nie je schopný ovládať alebo prežiť orgazmus.

Niektoré typy dysfunkcie sa vyskytujú u mužov aj u žien (napríklad nedostatok sexuálnej túžby). Ženy sa však častejšie sťažujú na subjektívnu kvalitu sexuálneho prežívania (napríklad chýba im radosť alebo záujem). Muži častejšie referujú o neschopnosti erekcie alebo ejakulácie aj napriek zachovanej sexuálnej apetencii. Sexuálnej dysfunkcii, najmä u mužov sa venuje čoraz väčšia pozornosť, čo sa odráža i vo vzrastajúcom počte epidemiologických štúdií. [1]

Kód F52 (sexuálna dysfunkcia) môže byť špecifikovaný – spresnený – ďalším číslom. Predkladáme prehľad sexuálnych neorganických sexuálnych dysfunkcií s krátkym popisom poruchy [11]. F52.0: Nedostatok alebo strata sexuálnej túžby. Nevylučuje schopnosť sexuálneho styku, ale znamená, že postihnutý pravdepodobne sexuálnu aktivitu nezačne. F52.1: Odpor k sexu(alite) a nedostatok sexuálnej slasti. Vyhliadky na sexuálnu interakciu s partnerom sú spojené so silnými negatívnymi pocitmi a vyvolávajú taký strach a úzkosť, že sa subjekt sexuálnej aktivity vyhne. F52.2: Zlyhanie genitálnej odpovede. U mužov je základným problémom dysfunkcia erekcie. Ak je erekcia v niektorých situáciách normálna, napríklad pri masturbácii, v spánku alebo s inou partnerkou, potom je príčina pravdepodobne psychogénna – neorganická. U žien je základným problémom suchosť vagíny a neschopnosť lubrikácie. F52.3: Dysfunkčný orgazmus. Orgazmus sa alebo vôbec nevyškytuje, alebo je veľmi opozdený. Táto porucha je častejšia u žien, u mužov sa vyskytuje málo. F52.4: Predčasná ejakulácia. Je to neschopnosť dostatočne ovládať ejakuláciu, aby pri sexuálnom styku dosiahli obidvaja partneri uspokojenie. F52.5: Neorganický vaginizmus. Kľč svalov panvového dna, ktoré obklopujú vagínu bráni otvoreniu vagíny a zavedeniu penisu je buď úplne nemožné alebo veľmi bolestivé. F52.6: Neorganická dyspareunia je bolesť pri sexuálnom styku. Ak nie je známa príčina tohto stavu, považujeme stav za psychogénny.

F52.7: Hypersexualita. V posledných rokoch sa táto diagnóza spochybňuje a takmer sa prestala používať, pretože odborníci sa nevedia zhodnúť na jednotnej, celosvetovo platnej definícii. Každá kultúra, každý národ má trochu iný vzťah k sexualite a preto dosiahnutie konsenzu pri tvorbe definície je v súčasnom svete takmer nemožné.

Sociálno-ekonomické faktory zohrávajú dôležitú úlohu v rozvoji sexuálnej dysfunkcie u oboch pohlaví. Španielski autori Fernandez Lozano a kol. [3] poukazujú na významné rozdiely vo vnímaní faktorov, ktoré prispievajú k sexuálnej spokojnosti v závislosti od veku, pohlavia, kultúrneho a ekonomického zázemia jedincov. Najčastejším sexuálnym problémom u mužov je erektilná dysfunkcia [6]. Incidencia sa zvyšuje s vekom. Porucha ovplyvňuje život až jednej tretiny mužov. Podstatne ovplyvňuje ich intímny život, kvalitu života a sebaúctu. Pri hodnotení sexuálnej dysfunkcie je nutné si uvedomiť, že bez ohľadu na príčinu (erektilnú a ejakulačnú) poruchu, vždy sú prítomné sekundárne psychické reakcie. Tieto podľa nich môžu zhoršiť sexuálnu dysfunkciu a/alebo ovplyvniť jej zvládanie, sexuálne vzťahy a compliance jedina pri terapii [10].

Podľa Kordu a kol. [8] prevalencia ženskej sexuálnej dysfunkcie je oveľa vyššia než prevalencia erektilnej dysfunkcie u mužov. Avšak len malá časť týchto žien vyhľadá lekársku pomoc a liečbu. Svoje tvrdenia opierajú o rozsiahlu prierezovú štúdiu. Ich cieľom bolo v nej zistiť prevalenciu a rizikové faktory sexuálnej dysfunkcie u nemeckých žien vo veku 20-80 rokov. Za týmto účelom zaslali 55-položkový dotazník, zahrňujúci FSFI (Female Sexual Function Index) 10 000 ženám (vek 20-80 rokov) v Kolíne nad Rýnom a okolia. Ich priemerný vek bol 43 rokov. Zo všetkých respondentiek (návratnosť dotazníkov bola 41%-ná) 38,2% malo sexuálnu dysfunkciu. Jej prevalencia výrazne vzrástla s vekom. Zatiaľ čo vo vekovej skupine <30 rokov 25,6% bolo postihnutých, v skupine 30-39 ročných 31,1%, v skupine 50-59 ročných až 44,8%. Vo vyšších vekových kategóriách sa sexuálna dysfunkcia vyskytovala v rozmedzí od 60-80% ($p=0,001$). Rizikovými faktormi boli vek, počet narodených detí, komorbidita (diabetes mellitus, vysoký krvný tlak, ochorenia srdca, ischemická choroba srdca, inkontinencia a depresia). Iba 18,4% týchto žien vyhľadalo lekársku pomoc. Z tohto výskumu vyplynulo aj to, že osoby (ženy dvakrát častejšie ako muži) s nižšími príjmami a jedinci (ženy aj muži v rovnakej miere) s nižším vzdelaním trpeli častejšie sexuálnou dysfunkciou.

Iní autori [7] zistili vo vzorke 454 respondentov, že u mužov, vysokoškolákov a jedincov s nízkou subjektívnou pohodou je častejšie prítomná sexuálna dysfunkcia, ktorú títo vnímajú ako obťažujúcu. Práve táto skutočnosť ovplyvňuje ich správanie pri hľadaní pomoci.

Giuliano a kol. [4] uskutočnili v rámci multicentrickej štúdie (27 krajín na celom svete) v európskom regióne na hodnotenie kvality sexuálneho života. krajín. V Európe bolo zapojených šesť krajín, a to Francúzsko, Taliansko, Nemecko, Španielsko, Švajčiarsko a Veľká Británia. V týchto krajinách uviedlo 40% - 82% mužov a žien, že nie sú úplne spokojní so svojim sexuálnym životom. Prítomnosť sexuálnej dysfunkcie nemusí nevyhnutne viesť k nespokojnosti so sexuálnym životom. [7] Nespokojnosť so sexuálnym životom súvisí so spokojnosťou so vzťahom, a so subjektívnou pohodou. Podľa zistení týchto autorov nesúvisí ani s demografickými charakteristikami.

Výskyt sexuálnych dysfunkcií je ťažké zisťovať, pretože získavanie seriózných údajov vo výskume je závislé od ochoty a spolupráce ľudí. Veľa osôb so sexuálnou dysfunkciou nikdy počas svojho života nevyhľadá lekársku pomoc, a to najmä príslušníci starších generácií. Populačné odhady prevalencie nie sú k dispozícii. Odhady prevalencie získavané z údajov primárnej zdravotnej starostlivosti, resp. klinických štúdií sú obvykle vyššie.

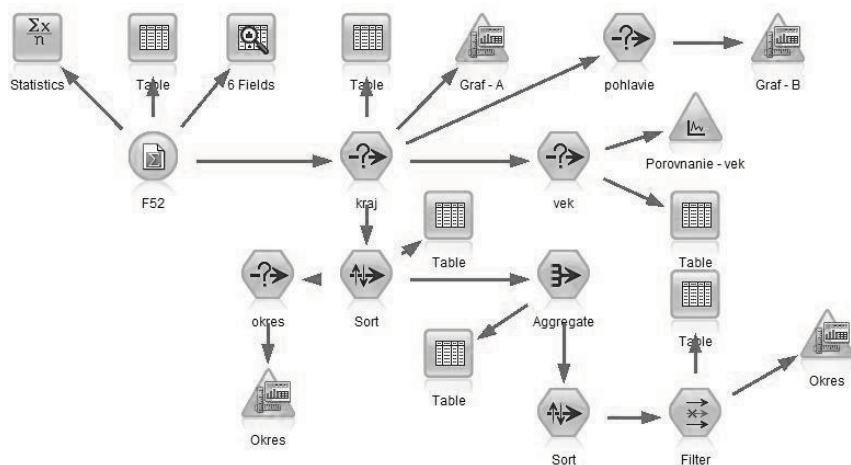
Presné odhady prevalencie, resp. incidence sú dôležité pre pochopenie skutočného zaťaženia mužov a žien sexuálnou dysfunkciou ako aj pre identifikáciu rizikových faktorov pre prevenciu. Cieľom nášho bádania bolo zistiť aká je výskyt a štruktúra diagnostikovaných

sexuálnych dysfunkcií (dg. F52) v Slovenskej republike a to podľa geografických celkov, pohlavia, veku a sociálnej skladby respondenta.

2. Materiál a metódy

Dáta boli získané z Národného centra zdravotníckych informácií SR za päťročné obdobie, a to od roku 2004 do roku 2008. Analýzy boli realizované pomocou metód dolovania dát. Dolovanie dát je proces objavovania zmysluplných nových korelácií, vzorov a trendov z veľkého množstva dát. Pri dolovaní dát sa používajú štatistické a matematické metódy. [9] Využitie metód dolovania dát je veľmi široké. Vo veľkom počte sa používa v ekonomike, pri zisťovaní podvodov, vo výrobnjej sfére, pri riadení procesov, predpovedaní vývoja sledovaného ukazovateľa a v neposlednom rade aj v zdravotníctve. [2], [13] Proces dolovania dát je súčasťou metodológie nazwanej Získavanie znalostí z databáz (Knowledge discovery in database – KDD). Veľmi dôležitým krokom KDD je príprava samotných dát, ktoré majú byť analyzované. Vo všeobecnosti sa môžeme stretnúť s dátami, ktoré nie sú homogénne alebo obsahujú chýbajúce hodnoty. Takéto dáta nemôžu byť pre následné analýzy nepresné a mohli by sme dostať skreslené alebo nepravdivé výsledky. Takisto je dôležité pripraviť dáta do takej podoby, aby mohli byť spracované vhodnými nástrojmi pre dolovanie dát. Na uchovanie veľkého množstva údajov je vhodné využiť dátové sklady, ktoré sú tiež súčasťou metodológie KDD. Dátové sklady poskytujú tzv. multidimenzionálny prístup k dátam. [5]

Ako prvotné dáta, ktoré sme spracovávali, boli údaje dodané Národným centrom zdravotníckych informácií SR za obdobie od roku 2004 do roku 2008. Tieto dáta boli vo forme Excelovskej tabuľky, čo bolo pre ďalšie analýzy nepostačujúce. Z pôvodnej Excelovskej tabuľky sme vyfiltrovali potrebné dáta, ktoré sme následne predpripravili do potrebného formátu pomocou nástroja SPSS Statistics. Následné analýzy sme vykonali pomocou nástroja SPSS Modeler, ktorý poskytuje širokú škálu matematických a štatistických metód. Obrázok 1 zobrazuje navrhnutý model, pomocou ktorého sme vykonali jednotlivé analýzy.



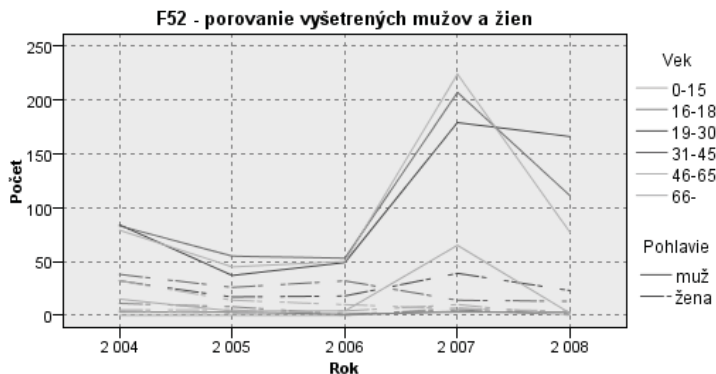
Obrázok 1: Navrhnutý model

Pre zobrazenie výsledkov sme využili deskriptívne nástroje, hlavne uzol Graphboard a Plot. Po nastavení správnych atribútov jednotlivých uzlov sme získali požadované výsledky.

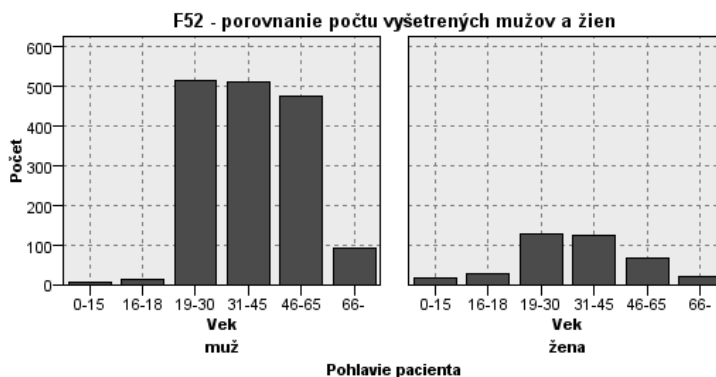
3. Výsledky a diskusia

Vo výsledkovej časti sme sa zamerali na nasledovné údaje: na zobrazenie vývoja počtu vyšetrených mužov a žien v jednotlivých vekových kategóriách, na porovnanie počtu vyšetrených mužov a žien, na sledovanie výskytu diagnózy v jednotlivých krajoch, na porovnanie počtu vyšetrených a hospitalizovaných.

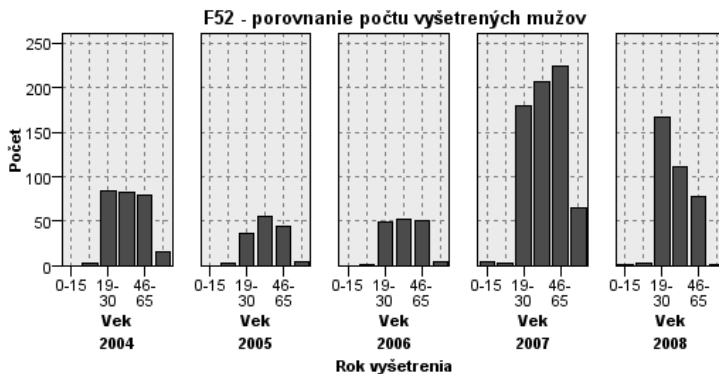
Počet prvých vyšetrení v ambulantnej starostlivosti v Slovenskej republike za päťročné obdobie, a to podľa pohlavia a veku sa rôzni. V roku 2007 bol zaznamenaný najvyšší počet prvých vyšetrení v celej Slovenskej republike, pričom v roku 2008 tento počet výrazne poklesol. (graf 1 a graf 3) U oboch pohlaví prevažovali jedinci vo vekovej kategórii 19-60 rokov, pričom vekom klesal počet vyšetrení. Ženy i v týchto vekových kategóriách sa výrazne zriedkavejšie obracali na odborníka pre diagnózu sexuálna dysfunkcia. (graf 2 a graf 4)



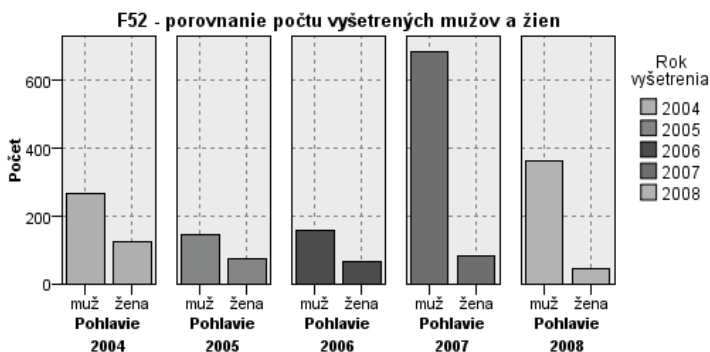
Graf 2: Porovnanie počtu vyšetrených mužov a žien na Slovensku v jednotlivých rokoch



Graf 2: Porovnanie počtu vyšetrených mužov a žien na Slovensku vo vekových kategóriách



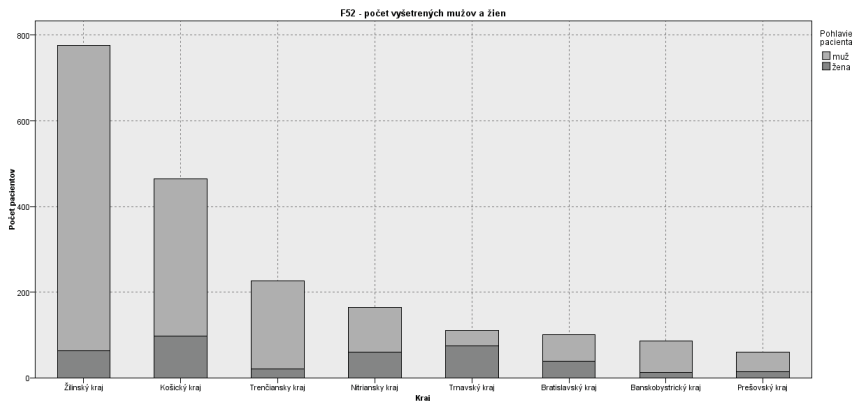
Graf 3: Porovnanie počtu vyšetrených mužov a žien



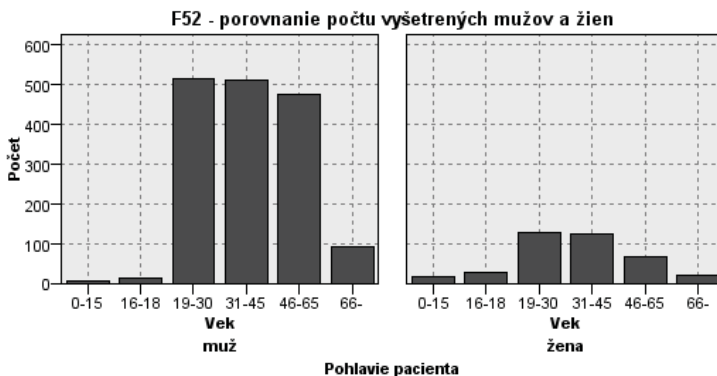
Graf 4: Porovnanie počtu vyšetrených mužov a žien

Najviac prých vyšetrení v ambulantej starostlivosti sa vyskytlo za sledované obdobie piatich rokov v Žilinskom kraji (graf 5) s výraznou prevahou mužov (graf 6) Najmenší počet vyšetrení u oboch pohlaví bol zaznamenaný v Prešovskom kraji.

Na diagnózu F52 bolo vyšetrených v Slovenskej republike 1989 osôb v ambulantej starostlivosti, z toho 1607 mužov a 382 žien. Na túto diagnózu boli hospitalizovaní len štyria pacienti za sledované päťročné obdobie. Ide o pacientov, ktorí sa vzdelaním pohybujú od základného neukončeného až po stredné s maturitou, nebývajú sami, sú slobodní.



Graf 5: Počet vyšetrených mužov a žien v jednotlivých krajoch



Graf 6: Porovnanie počtu vyšetrených mužov a žien v Žilinskom kraji

Na diagnózu F52 bolo vyšetrených v Slovenskej republike 1989 osôb v ambulantnej starostlivosti, z toho 1607 mužov a 382 žien. Na túto diagnózu boli hospitalizovaní len štyria pacienti za sledované päťročné obdobie. Ide o pacientov, ktorí sa vzdelaním pohybujú od základného neukončeného až po stredné s maturitou, nebývajú sami, sú slobodní.

4. Záver

Prevalenciu /incidenciu sexuálnych dysfunkcií je ťažké zisťovať v populácii. Väčšina postihnutých nikdy v živote nevyhľadá lekársku pomoc, pričom trpí nimi cca 19 – 50% žien a 50% mužov v závislosti od veku. V našej analýze sme zistili, že v sledovanom päťročnom období bol zaznamenaný najvyšší počet prvých vyšetrení v ambulantnej starostlivosti v roku 2007, a to vo všetkých sledovaných krajoch SR. Celkovo bol počet vyšetrení najvyšší za 5 ročné obdobie v Žilinskom kraji. Počet vyšetrení na diagnózu F-52 podľa pohlavia, veku a sociálnej skladby respondenta bol rozdielny. Údaje poukazujú na to, že mladšie ročníky vyhľadávajú pomoc sexuológa častejšie, a to predovšetkým v Bratislavskom kraji.

Tieto zistenia poukazujú na komplexný charakter problematiky, pričom viacerí odborníci indikujú na pasívny prístup k riešeniu problémov sexuálneho života, interpoplavné

rozdiely (4:1 v prospech mužov) vo vyhľadávaní odbornej pomoci a preferenciu anonymnej pomoci (internetové stránky, anonymné poradne). Výstupy analýzy poslúžia riadeniu odbornej starostlivosti na úrovni jednotlivých územných celkov pre rizikovú skupinu jedincov. Môžu poslúžiť zároveň pre návrh úpravy rozloženia poskytovania zdravotnej starostlivosti osobám so sexuálnymi poruchami.

5. Literatúra

- [1] BEUTEL, M. E., WEIDNER, W., BRÄHLER, E.: EPIDEMIOLOGY OF SEXUAL DYSFUNCTION IN THE MALE POPULATION. *ANDROLOGIA*, 2006, roč. 38, č. 4, s. 115-121. ISSN 0303456.
- [2] BOHÁČIK, J., DAVIS, D. N.: DATAMINING APPLIED TO CARDIOVASCULAR DATA. IN.: *JOURNAL OF INFORMATION TECHNOLOGIES*. 2010, roč. 3, č. 2, s. 14-21, ISSN 1337-7469.
- [3] FERNANDEZ LOZANO, A., SOTOLONGO VERGO, I., RUIZ CASTAÑE, E. FACTORS THAT CONTRIBUTE TO SEXUAL SATISFACTION. *J SEX MED*, 2007, č. 4 (SUPPL 2), s. 105. ISSN 1743-6109.
- [4] GIULIANO, F., JUENEMANN, K.-P., KIRBY, M., LEVINSON, I.: THE IMPORTANCE OF AND SATISFACTION WITH SEX IN EUROPE: REGIONAL ANALYSIS OF THE GLOBAL BETTER SEX SURVEY. THE 9TH CONGRESS OF THE EUROPEAN SOCIETY FOR SEXUAL MEDICINE VIENNA, AUSTRIA, DEC 3-6, 2006, *J. SEX. MED.* – 2007, č. 4, SUPPL 2, s. 150. ISSN 1743-6109.
- [5] HALENÁR, R.: ZAVÁDZANIE ÚDAJOV DO DÁTOVÉHO SKLADU A ICH TESTOVANIE. IN.: *JOURNAL OF INFORMATION TECHNOLOGIES*. roč. 2, č. 1, 2009, s. 7-14, ISSN 1337-7469.
- [6] HEIDELBAUGH, J. J.: MANAGEMENT OF ERECTILE DYSFUNCTION. *AM FAM PHYSICIAN*. 2010 FEB 1; roč. 81, č. 3, s. 305-312. ISSN 0002-838X.
- [7] KIRANA, P.-S., FRIDAKIS, M., MOISIDIS, K., ROSEN, R., MCKINLAY, J., HATZICHRISTOU, D.: FACTORS ASSOCIATED TO BOTHER FROM SEXUAL PROBLEMS: RESULTS FROM THE HOPPE'S STUDY. *J SEX MED*, 2007, č. 4, (SUPPL 2), s. 142. ISSN 1743-6109.
- [8] KORDA, J. B., BRAUN, M., CHRIST, H., HELLMICH, M., ENGELMANN, U.: PREVALENCE OF FEMALE SEXUAL DYSFUNCTION IN GERMANY: RESULTS OF THE CROSSOVER-STUDY "THE COLOGNE 20.000 COMMUNITY SURVEY". *J SEX MED*, 2007, č. 4 (SUPPL 2), s. 130. ISSN 1743-6109.
- [9] LAROSE, D.: DISCOVERING KNOWLEDGE IN DATA: AN INTRODUCTION TO DATA MINING. JOHN WILEY, 2005, 222 s., ISBN 0-471-66657-2.
- [10] LONGHI, E., SALONIA, A., COLOMBO, R., ZANNI, G., LANIA, C., BRIGANTI, A., RIGATTI, P., MONTORSI, F. THE ROLE OF THE SEXOLOGIST IN SEXUAL MEDICINE. *J SEX MEDICINE*, 2007, 4 (SUPPL 2), s. 105. ISSN 1743-6109.
- [11] MEZINÁRODNÍ KLASIFIKACE NEMOCÍ, 10. REVIZE. DUŠEVNÍ PORUCHY A PORUCHY CHOVÁNÍ. POPISY KLINICKÝCH PŘÍZNAKŮ A DIAGNOSTICKÁ VODÍTKA (PŘEL. Z ANGL. ORIG.). PRAHA, PSYCHIATRICKÉ CENTRUM 1992. 282 s – „ZPRÁVY“ č. 102. ISBN 80-85121-37-9.
- [12] MONTORSI, P., RAVAGNANI, P. M., GALLI, S., ROTATORI, F., BRIGANTI, A., SALONIA, A., DEHÒ, F., MONTORSI, F.: COMMON GROUNDS FOR ERECTILE DYSFUNCTION AND CORONARY ARTERY DISEASE. *CURR OPIN UROL*. 2004 Nov; roč. 14, č. 6, s. 61-65. ISSN 0963-0643.
- [13] TRNKA, A.: VYUŽITIE DOLOVANIA DÁT V ZDRAVOTNÍCTVE. IN.: *SLOVAK JOURNAL OF HEALTH SCIENCES*. 2010, roč. 1, č. 1, 2, s. 19-24, ISSN 1338-161X.

Adresa autorov:

Andrej Trnka, Ing. PhD.
Univerzita sv. Cyrila a Metoda v Trnave
Fakulta prírodných vied
Katedra aplikovanej informatiky
Nám. J. Herdu 2
917 01 Trnava
andrej.trnka@ucm.sk

Mária Kovářová, prof. MUDr. CSc.
Univerzita sv. Cyrila a Metoda v Trnave
Inštitút fyzioterapie, balneológie a liečebnej
rehabilitácie v Piešťanoch
Katedra klinických odborov a fyzioterapie
Rázusova 14
921 01 Piešťany
maria.kovarova@seznam.cz

Ivan Dóci, doc. MUDr. PhD.
Univerzitná nemocnica L. Pasteura
v Košiciach
II. psychiatrická klinika
Rastislavova 43
041 90 Košice
a
Univerzita Pavla Jozefa Šafárika v Košiciach
Lekárska fakulta
Trieda SNP 1
040 11 Košice
ivan.doci@upjs.sk

Porovnanie štatistickej gramotnosti krajín sveta

Comparison of Statistical Literacy in countries around the World

Eva Uhrinová

Abstract:

The article focuses on Statistical Literacy. We compare the level of Statistical Literacy of students in Slovakia with students from the other countries in the world. We mention projects dealing with the development of Statistical Literacy. We compare Statistical Literacy on the basis of research studies, namely PISA 2003, TIMSS 2003-2007. We comment upon the International Statistical Literacy Project, in which 66 countries are involved, including Slovakia.

Key words: Statistical Literacy, International Student Assessment, International Statistical Literacy Project

Kľúčové slová: štatistická gramotnosť, medzinárodné porovnávacie štúdie, Medzinárodný projekt štatistickej gramotnosti

JEL classification: C 18

1. Úvod

Význam štatistickej gramotnosti v dnešnej informačnej, dynamicky sa rozvíjajúcej spoločnosti stále narastá. V každodennom živote sa človek stretáva so štatistickými informáciami prezentovanými vo forme textov, čísel, percent, tabuliek, grafov. Schopnosť rozumieť týmto vyjadreniam informácií je preto podstatná. Túto skutočnosť si spoločnosť plne uvedomuje, a preto sa súčasťou učebných osnov základných aj stredných škôl stáva vyučovanie štatistiky. O úrovni štatistickej gramotnosti žiakov sa môžeme dozvedieť z medzinárodných meraní OECD PISA, alebo TIMSS .

V článku sa zaoberáme porovnaním štatistickej gramotnosti žiakov Slovenska so žiakmi z iných krajín sveta prostredníctvom výskumných štúdií PISA 2003, TIMSS 2003-2007. Pozornosť upriamujeme aj na Medzinárodný projekt štatistickej gramotnosti, do ktorého je zapojených 66 krajín sveta, vrátane Slovenska. Predostierame aj výsledky medzinárodnej súťaže v štatistickej gramotnosti, ktorú realizoval tento projekt, prostredníctvom ktorých chceme poukázať na krajiny, v ktorých majú žiaci dobré výsledky v štatistickej gramotnosti a ktorých školskú sústavu, učebné osnovy, či spôsob výučby je vhodné si preštudovať a nechať sa nimi inšpirovať pri výučbe našich žiakov.

2. Materiál a metódy

Štatistická gramotnosť spočíva v schopnosti narábať so štatistickými informáciami. Rozumieť im, vedieť ich interpretovať, vedieť ich ďalej používať.

Pojem štatistická gramotnosť sa vzťahuje na dve vzájomné zložky : (a) schopnosť ľudí interpretovať a kriticky hodnotiť štatistické informácie, údaje súvisiace s argumentovaním, alebo stochastickými javmi, s ktorými sa môžu stretnúť v rôznych kontextoch, (b) schopnosť diskutovať a komunikovať o štatistických údajoch, reagovať na štatistické informácie, ako aj chápanie významu týchto informácií [1].

Štatistickú gramotnosť žiakov Slovenska s krajinami sveta porovnáваме na základe výskumných štúdií z medzinárodných meraní PISA a TIMSS.

Štúdia PISA (Programme for International Student Assessment) je **medzinárodná porovnávací štúdia**, ktorá testuje úroveň vedomostí a zručností 15-ročných žiakov v oblasti matematickej, prírodovednej a čitateľskej gramotnosti v riešení problémov. PISA sa

zameriava na zisťovanie úrovne osvojenia si zručnosti, ktoré sú potrebné pre riešenie situácií z reálneho života. Dôraz sa kladie na uvažovanie, argumentáciu, komunikáciu, **orientáciu v grafoch a tabuľkách**, vyjadrenie bežných problémov v matematickom jazyku, riešenie problémov podľa návodu a **používanie štatistiky a pravdepodobnosti** [2].

V roku 2003 sa uskutočnil výskum PISA v 46 krajinách sveta. Do štúdie bolo zapojených 7500 15-ročných žiakov z 282 škôl všetkých typov a zo všetkých regiónov Slovenska. Tento reprezentatívny a stratifikovaný výber žiakov umožňuje zovšeobecniť získané výsledky na celý vzdelávací systém, a tak ho porovnať s výsledkami ostatných zúčastnených krajín [2].

Štúdia TIMSS (Trends in International Mathematics and Science Study, Trendy medzinárodnej štúdie matematiky a prírodovedných predmetov) je medzinárodná komparatívna štúdia, ktorá sa zameriava na zisťovanie vedomostí a zručností z matematiky a z prírodovedných predmetov u žiakov 4. a 8. ročníka základných škôl ako aj u žiakov končiacich strednú školu. Cieľom merania TIMSS je zistenie výkonov žiakov v medzinárodnom kontexte. Takýmto porovnaním výkonov získavajú zúčastnené krajiny informáciu o kvalite vlastného vzdelávacieho systému. Štúdia je organizovaná prostredníctvom Medzinárodnej asociácie pre hodnotenie výsledkov vzdelávania (IEA – International Association for the Evaluation of Educational Achievement) [3].

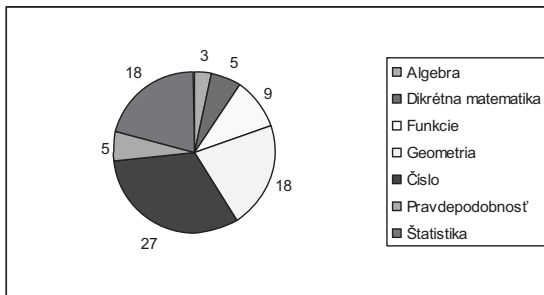
Medzinárodná štúdia TIMSS začala prebiehať v roku 1995. Slovensko sa zapojilo do štúdie TIMSS už v tomto prvom cykle. V TIMSS 1995, ale aj TIMSS-R 1999 a TIMSS 2003 sa testovania na Slovensku zúčastnili žiaci 8. ročníka základných škôl a 1. ročníka osemročných gymnázií. Žiaci 4. ročníka základných škôl boli do medzinárodného merania zapojení až v cykle TIMSS 2007 [3].

Do medzinárodného výskumu TIMSS 2003 sa celkovo zapojilo 48 krajín z celého sveta, s celkovo viac ako 395 000 žiakmi. Na Slovensku sa zapojili dve subpopulácie žiakov 8. ročníka základných škôl a žiaci 4. ročníka (kvarty) gymnázií s osemročným štúdiom.

3. Výsledky a diskusia

V štúdiu OECD PISA 2003 bola matematická gramotnosť zisťovaná prostredníctvom štyroch oblastí: 1. kvantita, 2. priestor a tvar, 3. zmena, vzťahy a závislosť, 4. **náhodnosť**.

Oblasť náhodnosť súvisí s našou problematikou, vychádza z **pravdepodobnosti a štatistiky**. Z tematického celku Pravdepodobnosť a štatistika bolo vo výskume PISA použitých celkom 23 úloh, z toho 5 bolo z pravdepodobnosti a 18 zo štatistiky (z celkových 85 matematických úloh) (obrázok 1). V porovnaní s našimi učebnými osnovami ide o zvýšenú pozornosť tomuto tematickému celku.



Obrázok 1: Počty úloh v jednotlivých oblastiach školskej matematiky v štúdiu PISA

Zo skúmaných oblastí **dopadla na Slovensku najhoršie oblast' náhodnosť**. Viac ako štvrtina žiakov **nedosahuje v tejto oblasti ani úroveň 2**. Táto úroveň je minimom, ktoré podľa autorov štúdie umožňuje samostatnú orientáciu a ďalšie celoživotné vzdelávanie žiaka.

V celkovej matematickej gramotnosti slovenský žiaci získali priemer 498 bodov (Hong Kong-550, Fínsko 544) a sme tak na 21. mieste spomedzi 46 zapojených krajín[4]. Porovnanie priemerného výkonu v oblasti Náhodnosť nám podáva nasledujúca tabuľka 1).

Tabuľka 1: Priemerný počet bodov vo vybraných krajinách v oblasti Náhodnosť [4]

p.č.	krajina	priemerný počet bodov	p.č.	krajina	priemerný počet bodov	p.č.	krajina	priemerný počet bodov
1.	Hong Kong -Čína	558	15.	Dánsko	516	29.	Portugalsko	471
2.	Holandsko	549	16.	Nórsko	513	30.	Taliansko	463
3.	Fínsko	545	17.	Švédsko	511	31.	Grécko	458
4.	Kanada	542	18.	Francúzsko	506	32.	Turecko	443
5.	Kórea	538	19.	Česká rep.	500	33.	Rusko	436
6.	Nový Zéland	532	20.	Rakúsko	494	34.	Srbsko	428
7.	Makao Čína	532	21.	Poľsko	494	35.	Thajsko	423
8.	Austrália	531	22.	Nemecko	493	36.	Uruguaj	419
9.	Japonsko	528	23.	Luxembursko	492	37.	Mexiko	390
10.	Island	528	24.	USA	491	38.	Indonézia	385
11.	Belgicko	526	25.	Maďarsko	489	39.	Brazília	377
12.	Lichtenštajnsko	523	26.	Španielsko	489	40.	Tunisko	363
13.	Írsko	517	27.	Slovensko	476			
14.	Švajčiarsko	517	28.	Lotyšsko	474			

Krajiny s poradovým číslo 1. až 17. sa nachádzajú štatisticky významne nad priemerom OECD, v krajinách s poradovým číslom 18. a 19. nie je štatisticky významný rozdiel vzhľadom na priemer OECD. Krajiny s poradovým číslom 20. a viac sa nachádzajú štatisticky významne pod priemerom OECD. **Slovensko** dosahuje priemerný počet bodov 476 a nachádza sa na škále Náhodnosti, štatisticky významne pod priemerom OECD.

Štúdie PISA 2006 a 2009 odhalili mierny, štatisticky zatiaľ nevýznamný, pokles priemerného výkonu OECD: o 2 body v roku 2006 (priemer OECD 498) a ďalšie 2 body v roku 2009 (priemer OECD 496) [5]. V týchto rokoch však neboli štúdie zamerané na matematickú gramotnosť (2006 – prírodovedná gramotnosť, 2009 – čitateľská gramotnosť). Najbližšie testovanie 15. ročných žiakov zameraných na matematickú gramotnosť sa uskutoční v roku 2012.

V medzinárodnom výskume TIMSS 2003 boli vedomosti a zručnosti z matematiky zisťované na základe piatich oblastí (tabuľka 2), z ktorých predmetom nášho výskumu je oblasť Údaje.

Tabuľka 2: Rozdelenie testových položiek z matematiky v štúdiu TIMSS

Výskumné oblasti obsahovej dimenzie	Počty položiek	%
Aritmetika	57	29,4
Algebra	47	24,2
Meranie	31	16,0
Geometria	31	16,0
Údaje	28	14,4
Spolu	194	100,0

Testové zošity obsahovali položky, ktoré experti IEA rozdelili do hlavných tematických okruhov v jednotlivých výskumných oblastiach. Hlavné tematické okruhy v rámci výskumnej

oblasti údaje boli nasledovné: 1. zhromažďovanie a triedenie údajov, 2. zobrazovanie údajov, 3. interpretácia údajov, 4. neurčitost' a pravdepodobnosť.

Medzinárodné priemerné skóre z matematiky dosiahlo hodnotu **467** bodov. Maximálne skóre z matematiky medzi zúčastnenými krajinami dosiahol Singapur 605 bodov a minimálne skóre Juhoafrická republika 264 bodov. Variáčne rozpätie skóre bolo 341 bodov.

Slovenská republika dosiahla priemerné skóre z matematiky **508** bodov (13. miesto), ktoré bolo **štatisticky významne vyššie ako medzinárodné priemerné skóre**. Minimálne bodové skóre SR malo hodnotu 243 bodov a maximálne bodové skóre bolo 778 bodov. Výsledok 26 krajín z matematiky bol štatisticky významne lepší ako medzinárodný priemer. Štatisticky významne vyššie skóre z matematiky ako **Slovenská republika** dosiahli z európskych krajín napr. Belgicko (flámska časť), Holandsko, Estónsko a Maďarsko. Vo všetkých výskumných obsahových oblastiach matematiky dosiahla Slovenská republika štatisticky významne vyššie skóre ako bol medzinárodný priemer. Ak porovnáme dosiahnuté priemerné skóre v jednotlivých výskumných oblastiach, **výskumná oblasť Údaje** dosiahla na Slovensku najhoršie výsledky (tabuľka 2).

Tabuľka 3: Priemerné skóre výskumných oblastí u slovenských žiakov v TIMMS

Výskumné oblasti	Priemerné skóre	miesto
Aritmetika	514	11.
Meranie	518	12.
Algebra	505	13.
Geometria	501	13.
Údaje	495	20.

Krajiny, ktoré v oblasti údaje dosiahli významne lepší výsledok ako medzinárodný priemer, mali priemerné skóre v intervale od 484 bodov (Rusko) do 579 bodov (Singapur). Variáčne rozpätie priemerného skóre medzi krajinami v oblasti údaje bolo 286 bodov.

Tabuľka 4: Priemerné skóre z oblasti údaje [6]

p.č.	krajina	Priemerné skóre	p.č.	krajina	Priemerné skóre	p.č.	krajina	Priemerné skóre
1.	Singapur	579	17.	Malajzia	505	32.	Macedónsko	419
2.	Japonsko	573	18.	Litva	502	33.	Indonézia	418
3.	Kórejská rep.	569	19.	Nórsko	498	34.	Bahrajn	414
4.	Čína - Taiwan	568	20.	Slovensko	495	35.	Čile	412
5.	Hongkong	566	21.	Slovensko	494	36.	Irán	404
6.	Holandsko	560	22.	Izrael	492	37.	Libanon	394
7.	Belgicko	546	23.	Taliano	490	38.	Egypt	393
8.	Švédsko	539	24.	Rusko	484	39.	Palestína	390
9.	Estónsko	535	-	Medzin. priemer	467	40.	Filipíny	390
10.	Anglicko	535	25.	Bulharsko	458	41.	Tunisko	387
11.	Austrália	531	26.	Cyprus	458	42.	Botswana	375
12.	Škótsko	531	27.	Srbsko a ČH	456	43.	Maroko	374
13.	USA	527	28.	Rumunsko	445	44.	Saud. Arábia	339
14.	Maďarsko	526	29.	Jordánsko	430	45.	JAR	296
15.	Nový Zéland	526	30.	Moldavsko	428	46.	Ghana	293
16.	Lotyšsko	506	31.	Arménsko	419			

Dosiahnuté priemerné skóre slovenských žiakov z matematiky pokleslo z hodnoty 534 bodov v roku 1995 na hodnotu 508 bodov v roku 2003, zistili sme štatisticky významný pokles vzdelávacích výsledkov.

Cieľová populáciu v štúdiu TIMSS 2007 predstavujú (predovšetkým) **žiaci 4. ročníka**. Meranie výkonov žiakov v matematike vo 4. ročníku sa uskutočnilo v troch obsahových oblastiach: Čísla, Geometrické útvary a merania a **Zobrazovanie údajov**. Oblasť **Zobrazovanie údajov** bola zameraná na **čítanie a interpretáciu jednoduchých tabuliek**. V testoch bolo zastúpených najviac úloh z oblasti **Čísla** (cca 50 %) a najmenej z oblasti **Zobrazovanie údajov** (15 %). Z oblasti **Zobrazovanie údajov**, získalo Slovensko priemerné

skóre 492 (23. miesto), najlepšie skóre má Hongkong (585), Najnižšie Tunisko (307). Česká republika (493) a Maďarsko (504) majú o niečo vyššie skóre [7].

Medzinárodný projekt štatistickej gramotnosti

Štatistickú gramotnosť jednotlivých krajín vo svete, môžeme sčasti porovnávať aj prostredníctvom **Medzinárodného projektu štatistickej gramotnosti** -International Statistical Literacy Project. Poslaním Medzinárodného projektu štatistickej gramotnosti je podporovať, vytvárať, spolupodieľať sa na **aktivitách podporujúcich štatistickú gramotnosť** po celom svete. Je v ňom zapojených 66 krajín sveta, vrátane Slovenska. Na stránke tohto medzinárodného projektu [8] sú poskytované informácie o rôznych novinkách v štatistickej gramotnosti, informácie o medzinárodných aktivitách, ktoré majú slúžiť na zvýšenie štatistickej gramotnosti.

Zo Slovenska je v tomto projekte zapojená Veronika Uričová, výskumná pracovníčka národného ústavu certifikovaných meraní vzdelávania (Núcem), ktorá sa v súčasnosti venuje **výskumu intervencie na zvýšenie štatistickej a finančnej gramotnosti slovenských žiakov na úrovni ISCED2**.

V školskom roku 2008-2009 sa pod záštitou tohto projektu uskutočnila prvá **medzinárodná súťaž v štatistickej gramotnosti**. Zúčastnili sa jej krajiny: Južná Afrika, Portugalsko, Taliansko, Fínsko, USA, Nový Zéland, Čína, Kanada, Argentína, Brazília, Čile, Kolumbia, Španielsko, Holandsko, Mexiko, Francúzsko, Kostarika, Srí-Lanka, Indonézia, Nigéria, Filipíny, Rusko, Malawi, Lesotho, Veľká Británia. Slovensko sa tejto súťaže nezúčastnilo. Táto súťaž prebiehala v troch kolách – školskom, národnom, medzinárodnom. Medzinárodné kolo sa uskutočnilo v juhoafrickom Durbane v auguste 2009, zúčastnilo sa ho 44 detí.

Výsledky medzinárodnej štatistickej súťaže sú nasledovné [8]:

Veková kategória 12-14 rokov: 1. miesto Argentína, 2. miesto Kanada, 3. miesto Malawi

Veková kategória 15-16 rokov: 1. miesto Malawi, 2.miesto Južná Afrika, 3. miesto Taliansko

Veková kategória 17- 18 rokov: 1. miesto Fínsko, 2.miesto Lesotho, 3. miesto USA

V školskom roku 2010-2011 je vyhlásená ďalšia medzinárodná súťaž štatistickej gramotnosti **súťaž o najlepší plagát**. Súťaž je určená pre študentov základných a stredných škôl a jej cieľom je snaha pomôcť študentom so zlepšovaním zručností popísať svoje prostredie pomocou štatistiky, použiť štatistiku ako nástroj k hľadaniu zmysla v každodennom živote. Víťazné plagáty z národnej súťaže budú vystavené na Svetovom štatistickom kongrese v Dubline. Do tejto súťaže sa zapojili krajiny: Arménsko, Argentína, Austrália, Benin, Brazília, Bulharsko, Kamerun, Kanada, Kapverdy, Kolumbia, Ekvádor, Etiópia, Fínsko, Francúzsko, Veľká Británia, Maďarsko, Irán, Írsko, Taliansko, Japonsko, Mexiko, Mozambik, Nový Zéland, Pakistan, Portugalsko, Rumunsko, Senegal, Španielsko, Česká republika, Ukrajina, USA. Žiaľ, Slovensko sa ani tejto súťaže nezapojilo[8]

Slovensko sa nezapojilo ani do jednej z týchto súťaží, práve preto nevieme porovnať úroveň štatistickej gramotnosti skúmané touto súťažou Slovenska s ostatnými zapojenými krajinami, môžeme sa však nechať inšpirovať výsledkami ostatných krajín. Snáď väčšia informovanosť o tomto projekte a o potrebe zvyšovania štatistickej gramotnosti, zvýši ochotu zapájania sa do medzinárodných súťaží, ktoré sú priamo zamerané na zvyšovanie štatistickej gramotnosti.

4. Záver

Z oboch spomínaných medzinárodných štúdií, PISA a TIMMS, vieme vyvodit' závery o stave štatistickej gramotnosti slovenských žiakov a jeho porovnaní s inými krajinami sveta, ktoré sa zapojili do výskumných štúdií, v PISA skúmaním výskunnej oblasti Náhodnosť a v TIMMS výskumnou oblasťou Údaje, resp. Zobrazovanie údajov. V porovnaní s ostatnými výskunnými oblasťami, tieto oblasti dopadli na Slovensku najhoršie.

V PISA 2003 dosahuje Slovensko 27. miesto na škále Náhodnosti (priemerné skóre 476) a nachádza sa štatisticky významne pod priemerom OECD. V TIMMS 2003 Slovenská republika dosiahla priemerné skóre z výskumnej oblasti Údaje 495 bodov (20. miesto), čo je štatisticky významne vyššie skóre ako bol medzinárodný priemer. V TIMMS 2007 v oblasti zobrazovanie údajov, získalo Slovensko priemerné skóre 492 (23. miesto), nachádzame sa na úrovni priemeru TIMSS 2007, ale na úrovni pod priemerom krajín OECD aj krajín EÚ zapojených do tohto merania.

Krajiny, ktoré dosahujú popredné pozície v medzinárodných štúdiách, v oblastiach štatistickej gramotnosti, či medzinárodných súťažiach zameraných na štatistickú gramotnosť, môžu byť pre ostatné krajiny príkladom v ich školskej sústave, učebných osnovách, či spôsobe výučby. Zapájaním žiakov do súťaží štatistickej gramotnosti, zvýšime záujem žiakov o túto dôležitú časť matematiky, potrebnú pre život človeka v dnešnej modernej spoločnosti.

5. Literatúra

- [1] INTERNATIONAL STATISTICAL REVIEW, 2002. Volume 70, číslo 1, [2011-04-06]. Dostupné na internete: <<http://www.stat.auckland.ac.nz/~iase/cblumberg/gal.pdf>>
- [2] MATEMATICKÁ GRAMOTNOSŤ VO VÝSKUME PISA 2003. [2010-02-26]. Dostupné na internete: <<http://www.ineko.sk/ostatne/matematicka-gramotnost>>
- [3] JELEMENSKÁ, J. – ĎURĎOVIČOVÁ, L. 2008. *Zbierka z uvoľnených úloh medzinárodných meraní TIMSS*, 1. vyd. Bratislava: ŠPÚ, 2008. 103 s. ISBN 978-80-89225-43-9 [2010-04-08] Dostupné na internete: <http://www.nucem.sk/documents//27/medzinarodne_Merania/timss/publikacie/zbierka_web.pdf>
- [4] KUBÁČEK, Z. ai. 2004 *Matematická gramotnosť*. Bratislava: ŠPÚ [2010-02-26] Dostupné na internete: <http://www.vyskummladeze.sk/images/stories/iuventa/DAVM_009/Pisa_mat_gram_sprava.pdf> ISBN 80-85756-88-9
- [5] NÁRODNÁ SPRÁVA PISA 2009 SLOVENSKO, Bratislava: NÚCEM BA [2010-02-26] Dostupné na internete: http://www.nucem.sk/documents/27/medzinarodne_merania/pisa/publikacie/Narodna_sprava_Slovensko_PISA_2009.pdf
- [6] KURAJ, J. – KURAJOVÁ, J. 2003. *Národná správa TIMSS 2003*. 1. vyd. Bratislava: ŠPÚ, 2006. 250 s. ISBN 80-89225-22-5 [2011-04-06]. Dostupné na internete: <http://www.nucem.sk/documents//27/medzinarodne_merania/timss/publikacie/Kuraj-Stopkova_Narodna_sprava_TIMSS2003.pdf>
- [7] JELEMENSKÁ, P. 2008. *Výkony žiakov 4. ročníka ZŠ v matematike. Národná správa z merania TIMSS 2007*. 1. vyd. Bratislava: ŠPÚ, 2008. 47 s., ISBN 978-80-89225-44-6 [2011-04-06]. Dostupné na internete: http://www.nucem.sk/documents//27/medzinarodne_merania/timss/publikacie/N%C3%A1rodn%C3%A1_spr%C3%A1va_web.pdf
- [8] **Medzinárodný projekt štatistickej gramotnosti**, [2011-04-09]. Dostupné na internete: <<http://www.stat.auckland.ac.nz/~iase/islp/home>>

Adresa autora (-ov):

Eva Uhrinová, Mgr.
Katedra matematiky UKF Nitra
Trieda A. Hlinku 1
94974 Nitra
eva.uhrinova@ukf.sk

Pravdepodobnostné rozdelenie likvidného rizika The Probability Distribution of The Liquidity Risk

Marta Vrábelová, Marta Urbaníková

Abstract: The aim of this paper is to estimate the two dimensional probability distribution of random vector with two indexes of liquidity as components from the data of Central and Eastern Europe banks.

Key words: Liquidity risk, Weibull distribution, normal distribution, Frank copula.

Kľúčové slová: Likvidné riziko, Weibullovo rozdelenie, normálne rozdelenie, Frankova kopula.

JEL classification: C81, C16

1. Úvod

Likvidita je schopnosť banky vyplatiť svoje záväzky bez neprijateľných strát, teda mať dostatok hotovosti alebo aktív, ktoré sa dajú v prípade potreby použiť. V tomto príspevku sledujeme dva ukazovatele likvidného rizika a našim cieľom je odhadnúť tak pravdepodobnostné rozdelenie jednotlivých ukazovateľov likvidného rizika ako aj vektora vytvoreného z daných ukazovateľov. Pravdepodobnostné rozdelenie odhadujeme na základe údajov získaných zo 106 bánk. Na modelovanie marginálnych pravdepodobnostných rozdelení používame Weibullovo a normálne rozdelenie, dvojrozmerné rozdelenie sme odhadli ako súčin marginálnych hustôt a tiež použitím Frankovej kopule. Parametre marginálnych rozdelení sme odhadli zvlášť pomocou štatistického softvéru a potom sme odhadli parameter závislosti metódou maximálnej vierohodnosti. Ukazovatele likvidity boli použité aj v článku [2] pri porovnávaní bánk krajín strednej a východnej Európy. V článku ide predovšetkým o poukázanie na možnosti a na postup pri odhade dvojrozmerných pravdepodobnostných rozdelení. Na výpočty sme použili program Mathematica a štatistický softvér SPSS. Základným prameňom k problematike kopúl je Nelsenova kniha [4], pre aplikácie možno odporučiť publikáciu [5].

2. Použité údaje

Pri skúmaní likvidného rizika sme použili údaje o bankách z rokov 2006-2008 u takmer všetkých bánk strednej a východnej Európy z databázy Bankscope a z výročných správ jednotlivých bánk. Sledovali sme tieto ukazovatele likvidného rizika:

- podiel hotovosti a obchodovateľných cenných papierov k celkovým aktívam (Cash and Due from Banks/ Total Assets)
- podiel likvidných aktív so splatnosťou do 3 mesiacov k likvidným pasívam so splatnosťou do 3 mesiacov (Liquid Assets / Liquid Liabilities)

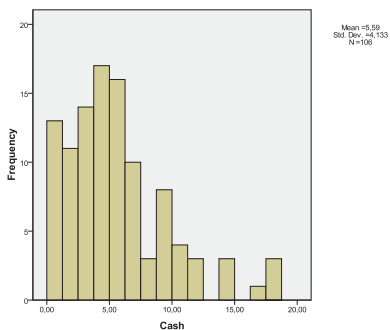
Prvý ukazovateľ označujeme ako premennú *Cash* a druhý ukazovateľ ako premennú *Liquid*. Vylúčili sme všetky rumunské banky a tiež 5 ďalších bánk, pretože vykazovali odľahlé hodnoty niektorého z ukazovateľov. Údaje o počte bánk sú uvedené v tabuľke 1. Popisné štatistiky skúmaných premenných obsahuje tabuľka 2 a ich histogramy sú na obrázku 1.

Tabuľka 1: Počet bánk krajín strednej a východnej Európy

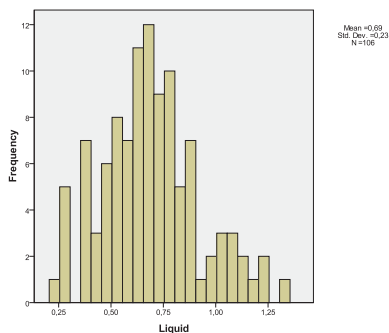
Krajina	Počet vybratých bánk	Počet vylúčených bánk
Lotyšsko	19	-
Estónsko	5	-
Litva	7	1
Poľsko	14	-
Maďarsko	8	1
Bulharsko	15	2
Slovensko	13	-
Slovenská republika	14	-
Česká republika	11	1
Spolu	106	5

Tabuľka 2: Popisné charakteristiky premenných Cash a Liquid

Descriptive Statistics			
	N	Mean	Std. Deviation
Cash	106	5,5898	4,13289
Liquid	106	,6858	,22963
Valid N (listwise)	106		



a)



b)

Obrazok 1: Histogram: a) premennej Cash, b) premennej Liquid

Normalitu rozdelenia premennej *Cash* zamietame, p-hodnota testov normality je 0,000 (tab. 3), u premennej *Liquid* normalitu rozdelenia nezamietame.

Koeficient korelácie premenných *Cash* a *Liquid* je rovný 0,049, čo svedčí o štatistickej lineárnej nezávislosti premenných.

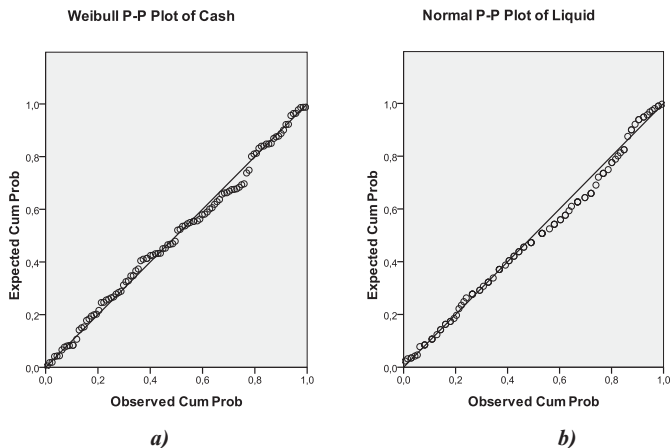
Tabuľka 3: Výsledky testov normality

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Cash	,131	106	,000	,898	106	,000
Liquid	,077	106	,141	,982	106	,171

a. Lilliefors Significance Correction

3. Marginálne rozdelenia

Rozdelenie premenných *Cash* a *Liquid* sme odhadli graficky v programe SPSS, využili sme P-P plot. Zo všetkých programom ponúkaných možností sme vybrali najvhodnejšie rozdelenie, ktorým bolo Weibullovo rozdelenie pre premennú *Cash* a normálne rozdelenie pre premennú *Liquid*. Pravdepodobnostné grafy (P-P plot) pre obidve premenné sú na obrázku 2.



Obrázok 2: Pravdepodobnostný graf: a) premennej *Cash*, b) premennej *Liquid*

Odhady parametrov Weibullovhého rozdelenia premenných *Cash* a *Liquid* sú uvedené v tabuľke 3.

Tabuľka 3: Odhad parametrov rozdelenia

a) pre premennú *Cash*

b) pre premennú *Liquid*

Estimated Distribution Parameters			Estimated Distribution Parameters		
		Cash			Liquid
Weibull Distribution	Scale	6,150	Normal Distribution	Location	,6858
	Shape	1,363		Scale	,22963

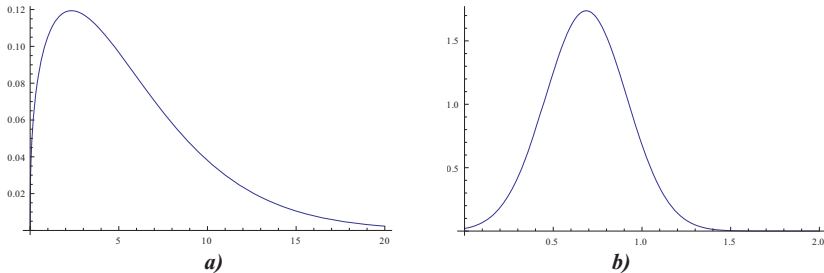
Hustotu rozdelenia pravdepodobnosti premennej *Cash* sme teda odhadli funkciou

$$f(x) = 0.11462e^{-0.0840937x^{1.363}}x^{0.363}$$

a hustotu premennej *Liquid* funkciou

$$g(y) = 1.73733e^{-9.48228(y-0.6858)^2}.$$

Grafy hustôt sú na obrázku 3.



Obrázok 3: Graf hustoty: a) premennej *Cash*, b) premennej *Liquid*

Hodnoty premennej *Cash* sme roztriedili do 10 intervalov a vypočítali sme hodnotu χ^2 -testu dobrej zhody s Weibulloým rozdelením, ktorá je rovná 6.26798. Štyri teoretické početnosti boli menšie ako päť, preto sme vypočítali p-hodnotu pre $\chi^2(5)$ rozdelenie v bode 6.26798, je rovná 0.28101. Lepšia zhoda rozdelenia s teoretickým Weibulloým rozdelením sa prejavila u premennej *Cash*, ale zhodu rozdelenia s teoretickým normálnym rozdelením nezamietame ani u premennej *Liquid*, p-hodnota pre Shapirov-Wilkov test z tabuľky 3 je rovná 0.171.

4. Dvojrozmerné rozdelenie

Vzhľadom na to, že koeficient korelácie je nevýznamný, mohli by sme premenné *Cash* a *Liquid* považovať za nezávislé (i keď z nekorelovanosti premenných v prípade nenormálneho rozdelenia nevyplýva ich nezávislosť) a vypočítať hustotu dvojrozmerného rozdelenia ako súčin marginálnych hustôt

$$h(x, y) = 0.199132e^{-0.0840937x^{1.363}-9.48228(y-0.6858)^2}x^{0.363}.$$

Jej graf je na obrázku 4.

V prípade, že pripustíme závislosť premenných, tak by sme na odhad dvojrozmerného rozdelenia mali použiť niektorú kopulu. Použili sme Frankovu kopulu, ktorá patrí do skupiny archimedovských kopúl. Je to funkcia $C: [0,1 \times 0,1] \rightarrow [0,1]$, pre ktorú platí

$$C(F(x), G(y)) = H(x, y),$$

kde F a G sú marginálne distribučné funkcie a H je združená distribučná funkcia náhodného vektora X, Y^T . Jej zmiešaná parciálna derivácia podľa u_1 a u_2 má tvar

$$c(u_1, u_2, t) = \frac{t(1-e^{-t})e^{-(u_1+u_2)}}{\left(\left(1-e^{-t}\right) - \left(1-e^{-tu_1}\right)\left(1-e^{-tu_2}\right)\right)^2}$$

a združenou hustotou náhodného vektora X, Y^T je potom funkcia

$$c F(x), G(y), t f(x)g(y).$$

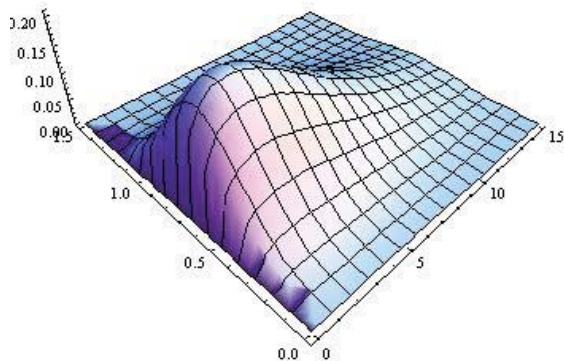
Parameter závislosti t sme odhadli metódou maximálnej vierohodnosti. Za u_1, u_2 sme dosadili hodnoty empirických distribučných funkcií F a G premenných *Cash* a *Liquid*, ktoré sme vypočítali z výberového súboru. Logaritmicou funkciou vierohodnosti je funkcia

$$\sum_{i=1}^{106} \ln c F(x_i), G(y_i), t + \sum_{i=1}^{106} \ln f(x_i)g(y_i) .$$

Parametre marginálnych rozdelení sme už odhadli, preto druhý výraz považujeme za konštantu. Prvý výraz je funkciou parametra t , maximum nadobúda v bode $t = 0.324658$. Hodnota Akaikeovho kritéria (bez druhého výrazu) je rovná

$$AIC = -2 \sum_{i=1}^{106} \ln c F(x_i), G(y_i), 0.324658 + 1 = 0.669875.$$

Čím je hodnota tohto kritéria nižšia, tým je výber kopule vhodnejší. Parameter Gaussovej, Studentovej, ani žiadnej inej archimedovskej kopule sa nám v programe Mathematica nepodarilo vypočítať, preto túto hodnotu nemáme s čím porovnať.



Obrázok 4: Graf hustoty dvojrozmerného rozdelenia

Hustota dvojrozmerného rozdelenia náhodného vektora $(X, Y)^T$ vypočítaná ako súčin marginálnych hustôt a hustoty kopule má tvar

$$\frac{0.0179226 e^{-0.0840937 x^{1.363} - 9.48228 (-0.6858-y)^2 - 0.324658 \left(1 - e^{-0.0840937 x^{1.363} + \frac{1}{2} (1 + \text{Erf}[3.07993 (-0.6858-y)])}\right)}}{\left(0.277226 - \left(1 - e^{-0.324658 \left(1 - e^{-0.0840937 x^{1.363}}\right)}\right)\right) \left(1 - e^{-0.162329 (1 + \text{Erf}[3.07993 (-0.6858-y)])}\right)}^2 x^{0.363}$$

Jej graf je však takmer identický s grafom na obrázku 4.

6. Záver

Marginálne rozdelenia pravdepodobnosti sledovaných bankových ukazovateľov likvidity sme odhadli Weibbulovým a normálnym rozdelením. Rozdelenie dvojrozmerného náhodného vektora ukazovateľov likvidity možno odhadnúť dvojrozmerným rozdelením, ktorého hustota je súčinom marginálnych hustôt, pretože možno predpokladať, že premenné *Cash* a *Liquid* sú nezávislé. Za predpokladu ich závislosti sme hustotu dvojrozmerného náhodného vektora odhadli aj s využitím Frankovej kopule a zistili sme, že grafy týchto dvoch hustôt sú takmer identické. Treba podotknúť, že presnejšie odhady parametrov by sme možno dostali, keby sme súčasne odhadovali parameter závislosti a parametre marginálnych rozdelení, čo sa nám v programe Mathematica nepodarilo urobiť.

7. Literatúra

- [1] ANDĚL, J. 2007. Základy matematické statistiky. MATFYZPRESS, Praha. ISBN 80-7378-001-1.
- [2] KÓŠOVÁ, M. 2010. Porovnanie krajín strednej a východnej Európy na základe hodnôt troch ukazovateľov stability banky. In: Forum Statisticum Slovaccum 5, s.115-120.
- [3] MEJSTRÍK, M – PEČENÁ, M. – TEPLÝ, P. 2008. Bankovníctví - Basic principles of banking. 1. vyd. Praha : Karolinum, 627 s. ISBN 978-80-246-1500-4.
- [4] NELSEN, R. 1999. An introduction to copulas. Springer New York.
- [5] TRIMEDI, P. K. – ZIMMER, D. M. 2005. Copula Modeling: An Introduction for Practitioners. In: Foundations and Trends in Econometrics, Vo. 1, No. 1, pp 1-111.

Adresy autorov:

Marta Vrábelová, doc. RNDr., CSc. KM FPV UKF, Tr. A. Hlinku 1 949 74 Nitra mvrabelova@ukf.sk	Marta Urbaníková, doc. RNDr., CSc. UMIT FPV UKF, Tr. A. Hlinku 1 949 74 Nitra murbanikova@ukf.sk
-------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------

Hodnotenie mikrobiologickej kvality vína využitím štatistických metód¹ Assessment of microbiological quality wines using statistical methods

Lucia Zeleňáková, Eva Matejková, Eubomír Lopašovský, Dagmar Kozelová

Abstract: Aim of this paper was to use statistical methods to assess the microbiological quality of wines in relation to the duration and conditions of their storage. We focused on the examination of factors (manufacturer, temperature and storage time) affecting the content of the wine yeast. The results showed that the way of storage wine in the room, resp. cooler temperature has significant influence on changes in the amount of yeast. It was confirmed that wine coming from the business network was better quality in contract to domestic wine.

Key words:

wine, microbiological quality, condition of storage, statistical methods

Kľúčové slová:

víno, mikrobiologická kvalita, podmienky skladovania, štatistické metódy

JEL classification:

C14, C20, L66

1. Úvod

V posledných rokoch spotreba vína stúpa, a s tým sa zvyšujú nároky konzumentov na kvalitu vín. Víno musí pôsobiť harmonicky a malo by vykazovať znaky špecifické pre danú odrodu[5]. Všeobecne je známe, že konzumácia vína v malom množstve môže pozitívne pôsobiť na ľudské zdravie [6].

Priaznivé účinky vína na ľudský organizmus sú známe už z obdobia 4000 rokov pred n.l. Možno ich zhrnúť do troch oblastí – antiklerotický efekt, antioxidačné účinky a antikarcinogénne účinky [2].

Víno patrí k obľúbeným nápojom spotrebiteľov o čom svedčí aj skutočnosť, že spotreba vína na obyvateľa v SR v roku 2009 vzrástla oproti predchádzajúcemu roku o 1,3 litra na 12,7 litra na obyvateľa. Produkcia hrozna a vinohradníctvo má na Slovensku dlhoročnú tradíciu. Rozloha rodiacich vinohradov v SR predstavuje podľa [7] plochu 15277,7 ha, nerodiace vinohrady do 3 rokov sa nachádzajú na výmere 661 ha a neobrábané vinohrady predstavujú 3695,3 ha. Vinohrady sa nachádzajú v 6 vinohradníckych oblastiach: Malokarpatská oblasť, Južnoslovenská, Nitrianska, Stredoslovenská, Východoslovenská a Tokajská vinohradnícka oblasť a celkovo predstavujú plochu 19634 ha. V sektore spracovania hrozna stále prebieha reštrukturalizácia s cieľom zvýšiť kvalitu vinohradníckej produkcie a konkurencieschopnosť domácej produkcie na medzinárodných trhoch.

Z technologického hľadiska je kvasenie hroznového muštu zložitý mikrobiologicko-biochemický proces, pri ktorom sa formuje kvalita dorábaného hroznového vína. Hroznový cukor v hroznovej šťave predstavuje nutričnú látku pre mnohé mikroorganizmy, hlavne pre kvasinky, v menšej miere pre baktérie. Mikroorganizmy v procese svojho metabolizmu premieňajú mušty na víno [1].

¹ Článok je súčasťou riešenia výskumného grantu KEGA 237-011SPU-4/2010 s názvom „Modernizácia a inovácia nových technológií vo výučbe a v špecializovaných laboratóriách hodnotenia kvality a bezpečnosti pokrmov“, ktorý sa rieši na Katedre hygieny a bezpečnosti potravín FBP SPU v Nitre

Kvasinky sú z vinársko-technologického hľadiska najdôležitejšie mikroorganizmy. Najviac kvasiniek sa nachádza na trhlínach bobúľ hrozna, kde presakuje hroznová šťava [9]. Výrobu vína v procese fermentácie a úlohu kvasiniek analyzovali tiež autori [3,4,8,10] a ďalší.

2. Cieľ, materiál a metódy

Cieľom príspevku je zhodnotiť mikrobiologickú kvalitu vybraných vín od 3 rôznych výrobcov vo vzťahu k dobe a podmienkam ich uchovávanía.

V práci sme sa v zmysle stanovených cieľov zamerali na skúmanie mikrobiologickej kvality vín (biele, suché, rizling vlašský) odoberaných počas 4 mesiacov od 3 rôznych výrobcov (po 10 fliaš od každého z nich). Vína od prvého a tretieho výrobcu boli odoberané z domácej výroby a víno od druhého výrobcu z obchodného reťazca. Sledovali sme vo víne prítomnosť kvasiniek, pričom sme laboratórnym skúšaním zisťovali ich počet v deň otvorenia fľaše (označenie 1), následne po 7 (označenie 2), 14 (označenie 3), 21 (označenie 4) a 28 (označenie 5) dňoch uchovávanía v chladničke a pri izbovej teplote.

V zmysle príslušných noriem pre stanovenie počtu kvasiniek sme na Petriho miskách odčítali vyrastené kolónie a prepočítali na konečnú hodnotu. Uvedené analýzy boli uskutočnené na Katedre hygieny a bezpečnosti potravín, FBP, SPU v Nitre. Za sledované obdobie sme spolu zanalyzovali 150 vzoriek.

Pri analýzach sme vychádzali z nasledovných pracovných hypotéz:

1. Víno odoberané z domácej výroby obsahuje vyšší počet kvasiniek na rozdiel od vína pochádzajúceho z obchodnej siete.
2. Počet kvasiniek postupne klesá počas skúmanej doby uchovávanía vína.
3. Teplota uskladnenia vína nemá výrazný vplyv na rast kvasiniek.

Z hľadiska štatistických metód boli pri analýzach použité nasledovné metódy:

- Jednofaktorová ANOVA, pričom rozdiely medzi úrovňami faktora boli posudzované na základe Scheffého testu.
- Podmienky homogenity rozptylov boli overované na základe Leveneho testu.
- V prípade nesplnenia podmienky homogenity bola použitá neparametrická obdoba ANOVY – Kruskal-Wallisov test.
- Index korelácie, ktorým sme posudzovali existenciu závislosti medzi obsahom kvasiniek vo víne a hodnotou pH vína.
- V aplikovaných analýzach bol použitý štatistický softvér Statgraphics.

3. Výsledky a diskusia

V zmysle stanovených cieľov sme sa zamerali na skúmanie vybraných faktorov determinujúcich kvalitu vína počas jeho uskladnenia v rôznych podmienkach.

Skúmali sme vplyv nasledovných podmienok:

1. Výrobca – ako bolo uvedené vyššie, hodnotili sme existenciu rozdielov z hľadiska obsahu kvasiniek vo víne medzi tromi výrobcami.
2. Uskladnenie – zisťovali sme obsah kvasiniek vo víne v dvoch podmienkach uskladnenia: pri izbovej teplote a v chladničke.
3. Doba uskladnenia – z hľadiska času skladovania vína sme robili merania piatich časových odstupoch, ktoré sú v analýzach označované ako spôsob 1, 2, 3, 4, 5.

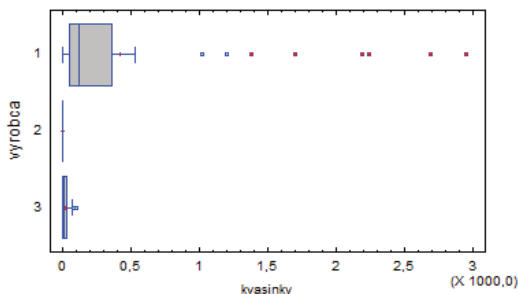
Spôsob 1: merania robené hneď po otvorení fľaše vína, Spôsob 2: merania robené hneď po 7 dňoch, Spôsob 3: merania robené hneď po 14 dňoch, Spôsob 4: merania robené hneď po 21 dňoch, Spôsob 5: merania robené hneď po 28 dňoch.

Hodnotenie kvality vína z hľadiska obsahu kvasiniek vo víne

Zmeny vo víne sú najčastejšie spôsobené prítomnosťou baktérií (mliečne, octové) a kvasinkových mikroorganizmov. Práve tie spôsobujú dodatočné kvasenie (refermentáciu) vín so zvyškom cukru a birzu. Všetky druhy kvasiniek a kvasinkových mikroorganizmov treba považovať za potenciálnych pôvodcov ochorenia a znehodnotenia sudových a fľaškových vín. I keď vývin octových baktérií vo víne je zriedkavejší, patria tieto mikroorganizmy k najnebezpečnejším nepriateľom zdravotného stavu vína.

V zmysle uvedeného sme v rámci mikrobiologickej analýzy skúmali vplyv vybraných faktorov (výrobca, podmienky uskladnenia, doba skladovania) na obsah kvasiniek vo víne. Pri skúmaní obsahu kvasiniek vo víne bol štatisticky vysoko preukazný vplyv výrobcu, t.j. boli potvrdené rozdiely v obsahu kvasiniek vo víne medzi jednotlivými výrobcami. Ako vyplýva aj z obrázku 1, výberový súbor *Výrobca 1* sa vyznačoval značnou heterogenitou, čo potvrdil aj Leveneho test pri overovaní podmienky použitia parametrickej ANOVY (p-hodnota Leveneho testu=0,00000278696). Vzhľadom na uvedenú skutočnosť sme testovali existenciu rozdielov pomocou neparametrickeho Kruskal-Wallisovho testu, ktorý potvrdil existenciu rozdielov (p-hodnota=0,0).

Obr. 1 Prezentovanie faktora VÝROBCA pomocou krabicových grafov



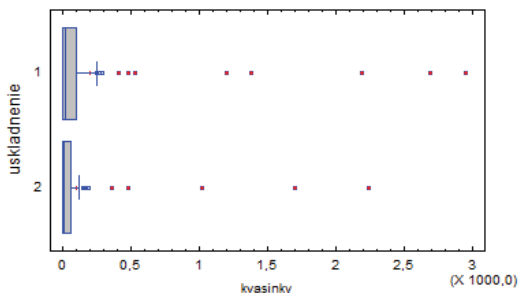
Pri zisťovaní existencie rozdielov medzi výrobcami boli vytvorené dve homogénne skupiny. Nižšie priemerné hodnoty kvasiniek sa nachádzali vo víne výrobcov 2 a 3, ktorí spolu tvorili spoločnú homogénnu skupinu. Oveľa vyšší priemerný obsah kvasiniek bol nameraný vo víne výrobcu 1, t.j. štatisticky významný rozdiel bol potvrdený medzi vínami tohto výrobcu v porovnaní s ostatnými výrobcami (tabuľka 1).

Tab. 1 Overovanie rozdielov medzi úrovňami faktora Výrobca pomocou Scheffeho testu

Výrobca	Počet	Priemer	Homogénne skupiny
2	50	0,0	X
3	50	22,56	X
1	50	421,88	X

Pri posudzovaní vplyvu druhého faktora – spôsobu uskladnenia vína – nebol potvrdený štatisticky významný vplyv daného faktora, t.j. nemôžeme zamietnuť hypotézu o neexistencii rozdielov medzi uskladnením vína pri izbovej teplote a uskladnením vína v chladničke, (p -hodnota=0,2080). Vychádzajúc z obrázku 2 sa zdá, že aj v tomto prípade sa výberové súbory javia ako heterogénne. Leveneho test však nevyvrátil podmienku o homogenite súborov (p -hodnota=0,208113).

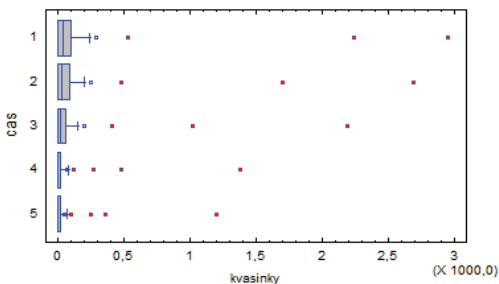
Obr. 2 *Prezentovanie faktora USKLADNENIE pomocou krabicových grafov*



Najpriaznivejšia skladovacia teplota pre biele vína je 10 až 12 °C a pre červené vína 12 až 14 °C. Ak víno nemožno zabezpečiť túto teplotu, musí sa počítať s tým, že proces jeho zrenia a starnutia plynie v takom prípade rýchlejšie ako normálne. Dôležité je, aby teplota bola stála a nekolísala. V chladných podmienkach víno dlho dokvasuje, pomaly sa čirí a dozrieva. V červených vínach dochádza k zrážaniu farbív. V príliš vlhkých pivniciach môže dôjsť k rozvoju nežiaducich baktérií a vláknitých mikroskopických húb. Relatívna vlhkosť pivnice má byť 70 až 80 %. Pri uskladňovaní vín nepriaznivo pôsobia vyššie teploty, kedy sa víno viac vyparuje, rýchlejšie dozrieva a starne. Pri vyšších teplotách sa mobilizuje činnosť nežiaducich mikroorganizmov (octové, mliečne baktérie) podstatne zhoršujúcich kvalitu vína a spôsobujúcich rôzne „ochorenia“ (octovanie, birzovatenie, vláčkovenie, manitové a mliečne kvasenie, maslové kvasenie, horknutie, zvrhnutie, myšina) a „chyby“ vín (sírovodík vo víne, hnednutie vína, príchut' po kvasniciach, kove, plesni, korku).

Pokiaľ ide o dobu uskladnenia vína (obrázok 3), závery sú identické s predchádzajúcim faktorom, tzn. aj doba uskladnenia štatisticky nevýznamne vplyva na obsah kvasiniek vo víne, (p -hodnota=0,5507).

Obr. 3 *Prezentovanie faktora ČAS pomocou krabicových grafov*



Legenda: čas: 1 – hneď po otvorení, 2 – po 7 dňoch, 3 – po 14 dňoch, 4 – po 21 dňoch, 5 – po 28 dňoch.

Analýzou sme zistili, že mierny pokles kvasiniek v závislosti od doby uchovávanía vína bol pravdepodobne spôsobený buď zvyškovým cukrom, z ktorého mikroorganizmy žili a neskôr odumierali alebo rozmnožením octových baktérií. Mikroorganizmy adaptované v prostredí, kde je zvyškový cukor, sú schopné sa množiť aj pri nižších teplotách, ale ich počet je o polovicu nižší ako pri teplotách vyšších. Mikrobiologické zmeny vo vínach boli zároveň sprevádzané senzorickými zmenami. Skladovaním vína v otvorených fľašiach dochádzalo k znižovaniu jeho kvality v zmysle výraznej zmeny vône a vzhľadu (octový zápach a zlomenie farby). Je vhodné „načaté“ víno čo najskôr vypiť, resp. opätovne uzatvoriť a skladovať v chlade. Niektoré mikroorganizmy potrebujú totiž na svoj rast a rozmnožovanie teplo, či prístup kyslíka.

V ďalšej časti analýz sme overovali existenciu závislosti medzi skúmanými premennými. Závislosť bola posudzovaná na základe indexov korelácie (tabuľka 2), ktoré boli odhadované na základe exponenciálnej, resp. polynomickej závislosti. V danej tabuľke sa nachádzajú hodnoty indexov korelácie v členení podľa podmienok uchovávanía, ako aj podľa doby skladovania. Výpočty sme robili spoločne pre výrobcu 1 a 3, pretože nebol zjavný rozdiel medzi nimi.

Tab. 2 Indexy korelácie: -pH- kvasinky v závislosti od doby ich uchovávanía

Podmienky uchovávanía	Doba skladovania (dni)				
	1 (0)	2 (7)	3 (14)	4 (21)	5 (28)
izbová teplota	0,94	0,91	0,86	0,96	0,96
chladnička	0,78	0,77	0,78	0,85	0,81

Ako vyplýva z tabuľky 2, vyššie závislosti medzi pH a obsahom kvasiniek vo víne sa dosahovali pri izbovej teplote. Pri uskladnení v chladničke môžeme pozorovať, že čím dlhšia doba skladovania vína, tým vyššia závislosť medzi pH a kvasinkami.

4. Záver

Jednou z najdôležitejších požiadaviek na kvalitu vína je jeho zdravotná bezpečnosť. Z vinársko-technologického hľadiska je mikrobiologická kvalita vína podmienená výsledkom vzájomného pôsobenia medzi kvasinkami, vláknitými mikroskopickými hubami a baktériami. Negatívne mikrobiologické vplyvy možno zabrániť najmä dokonalým zvládnutím výroby, pričom máme na mysli dostatočné odkalovanie muštov, správne zasienie, dokonalú filtráciu, maximálne dodržanie hygienických podmienok a pod.

Štatistickými metódami sme hodnotili vplyv vybraných faktorov (výrobca, teplota a dĺžka skladovania vína) na mikrobiologickú kvalitu vína (prítomnosť kvasiniek). Z výsledkov vyplynulo, že spôsob skladovania vína pri izbovej, resp. chladničkej teplote nemá preukazný vplyv na zmeny množstva kvasiniek. Naopak, potvrdilo sa, že najkvalitnejšie bolo víno pochádzajúce z obchodnej siete. Možno konštatovať, že z hľadiska kvality a bezpečnosti treba výrobe vín v domácich podmienkach venovať zvýšenú pozornosť.

Pre základ modernej vinárskej veľkovýroby sú nevyhnutné teoretické i praktické poznatky z oblasti chémie a mikrobiológie vína, rozšírené a doplnené poznatky z chémie muštu a vína, z fyziológie alkoholového a jablčno-mliečného kvasenia, z problematiky stabilizácie, filtrácie a kontroly vína, ako aj oblasti biochémie najdôležitejších mikroorganizmov vína. Domnievame sa, že poznatky i výsledky, ktoré ponúka táto práca, zohľadňujú súčasný stav i perspektívy vinárskej vedy, výskumu a praxe.

6. Literatúra

- [1] HRONSKÝ, Š. 2006. *Vinárstvo*. SPU: Nitra, 2006, 128 s. ISBN 80-8069-774-4.
- [2] CHLEBO, R. 2009. Víno a antioxidanty. In: KERESTEŠ, J. a i. 2009. *Biotechnológia, výživa a zdravie*. Považská Bystrica : Uniprint s.r.o. s. 347. ISBN 978-80-970205-9-0
- [3] LEJKOVÁ, L., JAVOREKOVÁ, S., KAČÁNIOVÁ, M., NOVÁKOVÁ, I., SELÉŠIOVÁ, Z., MAKOVÁ, J. 2011. Characteristic of microbiological community during fermentation of stum. In *Potravinárstvo*, roč. 5, 2011, p. 145 -150
- [4] NISIOTOU, A., NYCHAS, G. J. E. 2007. Yeast populations residing on healthy or *Botrytis*-infected grapes from a vineyard in Attica, Greece. In *Appl Environ Microbiol*. roč. 73, 2007. s. 2765–2768
- [5] SEDLO J., ŠEVČÍK J., LUDVÍKOVÁ I., 2004. Přehled odrůd révy 2004, 1. vyd. Velké Bílovice: Svaz vinařů České republiky, ISBN 80-9035-343-6. STEIDL, R., 2002. Sklepní hospodářství. 1. vyd. Valtice: Národní salón vín, 307 s. ISBN 80-903201-0-4.
- [6] STEIDL R., LEINDL G., 2004. Cesta ke špičkovému vínu, 1.vyd. Valtice: Národní salón vín, 67 s. ISBN 80-903201-4-7.
- [7] ŠAJBIDOROVÁ, V. 2010. Vinič hroznorodý, hroznové víno. Situačná a výhľadová správa k 31.7.2010. Bratislava : VÚEPP. s. 20. ISBN 978-80-8058-544-0
- [8] TOFALO, R., SCHIRONE, M., TELERA, G. C., MANETTA, A. CH., CORSETTI A., SUZZI, G. 2010. Influence of organic viticulture on non-*Saccharomyces* wine yeast populations In: *Annals of Microbiology*, Online First™, 24 August 2010
- [9] VAJČZIKOVÁ, I. – BREIEROVÁ, E. 2003. Identifikácia a druhové zastúpenie kvasiniek pri fermentácii hroznového muštu. In *Nova Biotechnologica*, roč.. 2, 2003, č. 3, s. 139 – 144. ISBN 80–89034–53–5.
- [10] VIANA, F, GIL, J. V., GENOVÉS, S., VALLÉS, S., MANZANARES, P. 2008. Rational selection of non-*Saccharomyces* wine yeasts for mixed starters based on ester formation and enological traits. In *Food Microbiol*. roč. 2008, č. 25, s.778–785

Kontakt:

Ing. Lucia Zelenáková, PhD.
Katedra hygieny a bezpečnosti potravín
Fakulta biotechnológií a potravinárstva
Slovenská poľnohospodárska univerzita
v Nitre
Trieda A. Hlinku 2, 949 76 Nitra
E-mail: lucia.zelenakova@uniag.sk

Ing. Eva Matejková, PhD.
Katedra štatistiky a operačného výskumu
Fakulta ekonomiky a manažmentu
Slovenská poľnohospodárska univerzita
v Nitre
Trieda A. Hlinku 2, 949 76 Nitra
E-mail: eva.matejkova@fem.uniag.sk

Ing. Dagmar Kozelová, PhD.
Katedra hygieny a bezpečnosti potravín
Fakulta biotechnológií a potravinárstva
Slovenská poľnohospodárska univerzita
v Nitre
Trieda A. Hlinku 2, 949 76 Nitra
E-mail: dagmar.kozelova@uniag.sk

MVDr. Lubomír Lopašovský, PhD.
Katedra hygieny a bezpečnosti potravín
Fakulta biotechnológií a potravinárstva
Slovenská poľnohospodárska univerzita
v Nitre
Trieda A. Hlinku 2, 949 76 Nitra
E-mail: lubomir.lopasovsky@uniag.sk

Hodnotenie kvality imunoenzymatických testov pri detekcii vzájomného falšovania kravského a ovčieho mlieka¹

Quality evaluation of immuno enzymatic tests for the detection of mutual adulteration cow and sheep milk

Lucia Zeleňáková, Ladislav Mura

Abstract: The main goal of this study was to evaluate the application of ELISA method (ELISA tests – RC bovine) for the detection of the lab-prepared adulteration of sheep milk with cow milk. The analyses were focused on laboratory testing and evaluation of specificity and sensitivity the ELISA tests. It was found out that the quality of adulterated milk detection is being impacted by a standard curve with a specific detection range. The samples pasteurized (71.7 - 77 °C for 20 sec.) in different combinations gave lower optical density responses than those prepared from the raw milk.

Key words: quality, detection of milk adulteration, calibration curve, regression

Kľúčové slová: kvalita, detekcia falšovania mlieka, kalibračná krivka, regresia

JEL classification:

C29, O13, Q16

1. Úvod

V súvislosti s platnosťou Potravinového kódexu SR musia požiadavky tuzemských spracovateľov mlieka, pokiaľ chcú uspieť na trhu, venovať maximálnu pozornosť kvalite svojich výrobkov, modernizovať výrobu i sortiment. Cieľom je výroba mliekarenských výrobkov s vyššími úžitkovými vlastnosťami, ktoré sú dané trvanlivosťou, balením, orientáciou na zvýraznenie výživových hodnôt a pod.

Množstvo mlieka u oviec sa mení počas dňa, v priebehu laktácie i s poradím laktácie. Na produkciu mlieka vplýva plemeno, rodičia, početnosť vrhu, výživa, klimatické podmienky, ako i mnohé ďalšie faktory. To nabáda spracovateľov ovčieho mlieka k jeho falšovaniu. Príčiny falšovania mlieka sú v prvom rade ekonomické dôvody [2]. Dôvody, prečo je odhalenie falšovania mlieka dôležité pre spotrebiteľa:

- pri miešaní drahšieho mlieka ovčieho a kozieho s kravským dochádza síce k redukcii výrobných nákladov, no zároveň aj k zníženiu produkcie kvalitných výrobkov.
- ochrana konzumentov, ktorí si neprajú konzumovať mlieko iných druhov. Isté percento obyvateľstva sa vyznačuje alergiou na zložky niektorých druhov mliek.

Monitoringu falšovania ovčieho mlieka a výrobkov z neho na spoločnom európskom trhu s potravinami sa vo svojom výskume venoval kolektív odborníkov [9]. Zistili, že najnovší trend u prvovýrobcov, ako aj spracovateľov ovčieho mlieka je znižovať legislatívou stanovené podiely ovčej hrudky na rozdiel od hrudky kravskej. Falšovanie mlieka a výrobkov z neho možno detegovať viacerými analytickými metódami. V potravinovej analýze je ELISA

¹ Článok je súčasťou riešenia výskumného grantu KEGA 237-011SPU-4/2010 s názvom „Modernizácia a inovácia nových technológií vo výučbe a v špecializovaných laboratóriách hodnotenia kvality a bezpečnosti pokrmov“, ktorý sa rieši na Katedre hygieny a bezpečnosti potravín FBP SPU v Nitre

najpoužívanejšia forma imunoanalýzy, pretože redukuje cenu vybavenia a je ľahko použiteľná, rýchla, pohotová a automatizovaná [1].

ELISA použili na detekciu kravského mlieka v ovčom mlieku a syroch aj iní [6]. Zistili vyššiu citlivosť tejto metódy na rozdiel od chromatografických metód. ELISA má výhody vo vysokej citlivosti, špecifickosti, nízkej cene a rýchlej aplikácii. Cieľovými antigénmi sú kazeíny, laktoglobulíny, imunoglobulíny alebo iné mliečne proteíny [5]. Nepriamu kompetitívnu ELISA metódu na detekciu nízkych koncentrácií nedeklarovaného kravského mlieka v kozom, ovčom a byvolom mlieku vyvinuli [4]. Uvedenú ELISA metódu použili aj [7]. Ďalší odborníci [3] použili sandwich ELISA na detekciu menej ako 1 % kozieho mlieka v ovčom mlieku. Výskumom potvrdili vysokú citlivosť uvedenej metódy. Napriek uvedeným výhodám, iba malý počet ELISA metód dokáže detegovať prídavky tepelne ošetreného mlieka [8].

2. Cieľ, materiál a metódy

Cieľom článku je na základe štatistických metód skúmať, či použité ELISA testy umožnia presnú a citlivú detekciu definovanej koncentrácie kravského mlieka v ovčom mlieku..

Analýzy sme uskutočnili v Národnom referenčnom laboratóriu pre mlieko a mliečne výrobky. Na vyšetrenie pripravených vzoriek mlieka sme použili komerčne ponúkaný ELISA test, pričom detekcia a kvantifikácia kravského mlieka bola založená na prítomnosti špecifických imunoglobulínov. Ovčie a kravského mlieko sme zmiešavali v stanovených pomeroch (0; 0,5; 1; 3; 10 a 15 % kravského mlieka v skúmanej vzorke mlieka), pričom sme vzorky podrobili v rôznych kombináciách tepelnému ošetrovaniu (71,7 – 77 °C po dobu 20 sekúnd). Vzniknuté zmesi sme laboratórne skúšali z rôznych hľadísk vybranými ELISA testami.

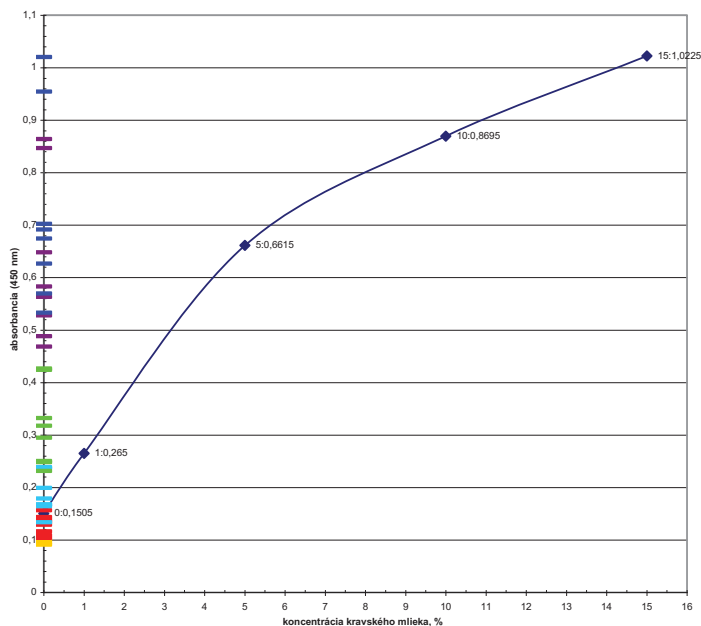
V rámci laboratórneho skúšania citlivosti a špecifickosti ELISA testov sme merania absorbancí sledovaných falšovaných podielov pri rôznych spôsoboch a rôznych pomeroch vzájomného zmiešavania opakovali dvakrát. Uvedenými testami sme pre hodnotenie sledovaných ukazovateľov analyzovali 43 vzoriek, ktoré zodpovedali 43 kombináciám zmiešavania kravského a ovčieho mlieka. Pre každý druh detekcie a pre každý test zvlášť sme uskutočnili 86 meraní. Pri hodnotení výsledkov sme aplikovali požiadavky výrobcov všetkých použitých testov, no z hľadiska objektívneho hodnotenia a porovnania kvality testov sme použili aj rovnaké hodnotiace princípy. Tie sa týkali zistenia koncentrácie falšovaného množstva mlieka iného druhu v skúmanej vzorke priamo interpoláciou nameranej absorbancie na kalibračnej krivke, ako aj matematickým výpočtom pomocou rovnice regresie, ktorá danej štandardnej krivke zodpovedá a umožňuje správnu determináciu koncentrácie falšovaného množstva.

Vzhľadom na široký rozsah výpočtov a na druhej strane limitovaný rozsah článku uvádzame iba vybrané výsledky nášho výskumu.

3. Výsledky a diskusia

Jednotlivé koncentrácie kravského mlieka v ovčom mlieku sme určili interpoláciou ich absorbancie na kalibračnej krivke (graf 1). Hoci výrobca testu RC – kravské odporúča aplikovať kalibračnú krivku s detekčným rozpätím 0 – 15 %, z hľadiska presnejšej determinácie skúmaných koncentrácií kravského mlieka v ovčom mlieku sme vytvorili niekoľko štandardných kriviek so špecifickým detekčným rozpätím. Každú štandardnú krivku sme následne doplnili trendovými čiarami lineárnej resp. polynomickej funkcie 2 stupňa. Pomocou regresných rovníc sme matematicky definovali priebeh jednotlivých funkcií a vypočítali presné koncentrácie podielov kravského mlieka v ovčom mlieku.

Graf 1 Kalibračná krivka na detekciu sledovaných koncentrácií kravského mlieka v ovčom mlieku pri rôznych spôsoboch vzájomného zmiešavania



Zdroj: vlastný výskum a spracovanie

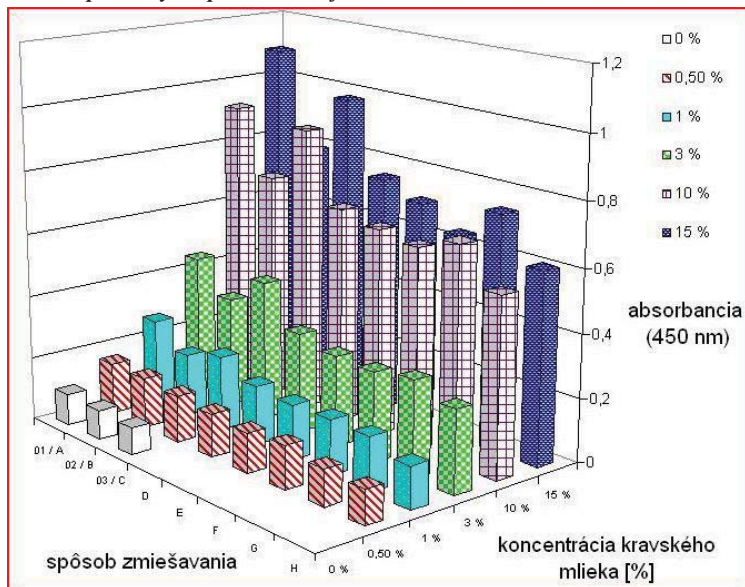
Výberom jednotlivých regresných modelov (lineárne a polynómické) sme pre jednotlivé detekčné rozpätia štandardných kriviek získali nasledovné alternatívne riešenia:

0 – 1 %	$y = 0,1145x + 0,1505$	$R^2 = 1$
0 – 5 %	$y = -0,0031x^2 + 0,1176x + 0,1505$	$R^2 = 1$
0 – 10 %	$y = 0,0723x + 0,1975$	$R^2 = 0,9548$
0 – 10 %	$y = -0,006x^2 + 0,133x + 0,1453$	$R^2 = 0,9997$
0 – 15 %	$y = -0,0036x^2 + 0,1105x + 0,1608$	$R^2 = 0,9946$ (odporúča výrobca testu)
5 – 10 %	$y = 0,0416x + 0,4535$	$R^2 = 1$
10 – 15 %	$y = 0,0306x + 0,5635$	$R^2 = 1$

Z hodnoty koeficientu determinácie (R^2) vyplýva, že uvedené funkcie vystihuje s veľmi vysokou tesnosťou. Porovnaním vypočítaných hodnôt absorbančie, resp. koncentrácií pomocou regresných rovníc možno konštatovať, že citlivosť, a teda aj kvalitu detekcie ovplyvňuje vytvorenie štandardnej krivky so špecifickým detekčným rozpätím. Uvedenú skutočnosť možno dokumentovať na nasledovnom príklade: na determináciu presnej koncentrácie kravského mlieka v ovčom mlieku pri nameranej absorbančii napr. 0,2385

možno použiť hneď niekoľko regresných rovníc. Najpresnejšiu hodnotu však možno dosiahnuť použitím rovnice zodpovedajúcej detekčnému rozpätiu 0 – 1 %.

Graf 2 RC test na detekciu falšovania sledovaných koncentrácií kravského mlieka v ovčom mlieku pri rôznych spôsoboch vzájomného zmiešavania



Zdroj: vlastný výskum a spracovanie

Keďže použitý ELISA test je určený na determináciu falšovaných podielov v surovom stave, zamerali sme sa najprv na hodnotenie kvality detekcie vzájomného zmiešavania surového kravského a surového ovčieho mlieka. Podľa výrobcu testu, vzorky vykazujúce tú istú alebo slabšiu absorbanciu ako 0 % štandard, neobsahujú kravské mlieko. Uvedenú skutočnosť sme zaznamenali aj v našom výskume, kde absorbancia vzorky bola 0,0995 oproti tej štandardnej s hodnotou 0,1505. Najväčší rozptyl hodnôt v rámci použitia rôznych detekčných rozpätí sme zaznamenali pri 3 % a 10 % podiele kravského mlieka v ovčom mlieku.

Zistili sme tiež, že 0,5 % podiel surového kravského mlieka test RC – kravské nezachytil v požadovanej miere (0,05 %). Naopak, najlepšiu detekčnú schopnosť preukázal test pri falšovaní ovčieho mlieka 10 % a 15 % podielom surového kravského mlieka (9,87 %, resp. 14,93 %). (tab. 1 a 2) Výsledky dosiahnuté pri ďalších koncentráciách dokumentujú početné tabuľky a grafy v rámci riešenia grantovej výskumnej úlohy, ale vzhľadom na obmedzený rozsah článku ich neuvádzame.

Tabuľka 1 RC test – kravské: porovnanie citlivosti detekcie falšovania kravským mliekom vybranými štandardnými krivkami s rôznym detekčným rozpätím

Koncentrácia kravského mlieka (%)	Spôsob zmiešavania	Absorbancia (450 nm)	Koncentrácia kravského mlieka v druhom mlieku vypočítaná pomocou regresných rovníc pri rôznych detekčných rozpätiach štandardných kriviek, %						
			0 – 1 %	0 – 5 %	0 – 10 %	0 – 10 %	0 – 15 %	5 – 10 %	10 – 15 %
			lineárna funkcia	polynomičná funkcia	lineárna funkcia	polynomičná funkcia	polynomičná funkcia	lineárna funkcia	lineárna funkcia
0	01	0,0995	•	•	•	•	•	•	•
	02	0,0945	•	•	•	•	•	•	•
	03	0,091	•	•	•	•	•	•	•
0,5	A	0,1565	0,05	0,05	*	0,09	*	————	————
	B	0,1435	•	•	•	•	•	•	•
	C	0,1385	•	•	•	•	•	•	•
	D	0,1345	•	•	•	•	•	•	•
	E	0,104	•	•	•	•	•	•	•
	F	0,129	•	•	•	•	•	•	•
	G	0,1165	•	•	•	•	•	•	•
	H	0,109	•	•	•	•	•	•	•
1	A	0,2385	0,77	0,76	0,57	0,72	0,72	————	————
	B	0,199	0,42	0,42	0,02	0,41	0,35	————	————
	C	0,237	0,76	0,75	0,55	0,71	0,71	————	————
	D	0,179	0,25	0,24	*	0,26	0,17	————	————
	E	0,164	0,12	0,12	*	0,14	0,03	————	————
	F	0,1675	0,15	0,15	*	0,17	0,06	————	————
	G	0,167	0,14	0,14	*	0,16	0,05	————	————
	H	0,1335	•	•	•	•	•	•	•

Zdroj: vlastný výskum a spracovanie

Tabuľka 2 RC test – kravské: pokračovanie

Koncentrácia kravského mlieka (%)	Spôsob zmiešavania	Absorbancia (450 nm)	Koncentrácia kravského mlieka v druhom mlieku vypočítaná pomocou regresných rovníc pri rôznych detekčných rozpätiach štandardných kriviek, %						
			0 – 1 %	0 – 5 %	0 – 10 %	0 – 10 %	0 – 15 %	5 – 10 %	10 – 15 %
			lineárna funkcia	polynomičná funkcia	lineárna funkcia	polynomičná funkcia	polynomičná funkcia	lineárna funkcia	lineárna funkcia
3	A	0,4265	————	2,51	3,17	2,37	2,63	————	————
	B	0,332	————	1,61	1,86	1,51	1,64	————	————
	C	0,424	————	2,49	3,13	2,34	2,6	————	————
	D	0,2945	————	1,27	1,34	1,19	1,26	————	————
	E	0,248	0,85	0,85	0,7	0,8	0,81	————	————
	F	0,2505	0,87	0,87	0,73	0,82	0,83	————	————
	G	0,3175	————	1,48	1,66	1,38	1,49	————	————
	H	0,2315	0,71	0,7	0,47	0,67	0,65	————	————
10	A	0,864	————	————	9,22	9,36	9,01	9,87	————
	B	0,648	————	4,85	6,23	4,83	5,34	————	————
	C	0,8465	————	————	8,98	8,64	8,63	9,45	————
	D	0,583	————	4,13	5,33	4,02	4,47	————	————
	E	0,528	————	3,54	4,57	3,4	3,79	————	————
	F	0,488	————	3,13	4,02	2,98	3,32	————	————
	G	0,563	————	3,91	5,06	3,79	4,22	————	————
	H	0,4685	————	2,93	3,75	2,78	3,1	————	————
15	A	1,0205	————	————	————	————	14,92	————	14,93
	B	0,7025	————	————	6,99	5,61	6,12	5,99	————
	C	0,9545	————	————	————	————	11,47	————	12,78
	D	0,6915	————	————	6,83	5,44	5,96	5,72	————
	E	0,6265	————	4,61	5,93	4,55	5,04	————	————
	F	0,5695	————	3,98	5,15	3,86	4,3	————	————
	G	0,6745	————	————	6,6	5,2	5,71	5,31	————
	H	0,533	————	3,59	4,64	3,45	3,85	————	————

Zdroj: vlastný výskum a spracovanie

4. Záver

V predkladanom článku sme pomocou štatistických metód skúmali, či použité ELISA testy umožnia presnú a citlivú detekciu definovanej koncentrácie kravského mlieka v ovčom mlieku. Detekcia a kvantifikácia kravského mlieka bola založená na prítomnosti špecifických imunoglobulínov. Z výsledkov vyplýva, že ELISA test nezachytil všetky sledované falšované koncentrácie na 100 %. Napriek tomu sme u sledovaných testov zaznamenali veľmi dobrú až výbornú schopnosť detegovať väčšinu sledovaných falšovaných prídavkov. Skúmaním vplyvu tepelného ošetrenia mlieka na detekciu falšovania sme zároveň zistili, že vzorky, ktoré boli v rôznych kombináciách tepelne ošetrené, vykazovali nižšie optické denzitné reakcie v porovnaní s tými, ktoré tepelnému ošetreniu neboli vôbec vystavené. Najvýraznejší pokles absorbancií sme zaznamenali pri zmiešavaní všetkých sledovaných pomerov pasterizovaného mlieka v pasterizovanom mlieku iného druhu – pasterizovaná zmes.

5. Literatúra

- [1] GIOVANNACCI, I. – GUIZARD, C. – CARLIER, M. et al. 2004. Species identification of meat products by ELISA. In *Int. J. Food Sci. Tech.*, 39, 2004, p. 863 –867.
- [2] HALKEN, S. 2003. Early sensitisation and development of allergic airway disease – risk factors and predictors. In *Paediatr. Respir. Rev.*, 4, p. 128 – 134.
- [3] HAZA, A.I. – MORALES, P. – MARTÍN, R. et al. 1997. Use of a monoclonal antibody and two enzyme-linked immunosorbent assay formats for detection and quantification of the substitution of caprine milk for ovine milk. In *Journal of Food Protection*, 60, 1997, p. 973 – 977.
- [4] HURLEY, I. P. – COLEMAN, R. C. – IRELAND, H. E.– WILLIAMS, S. 2004. Measurement of Bovine IgG by Indirect Competitive ELISA as a Means of Detecting Milk Adulteration. In *J. Dairy Sci.*, 87, p. 543 – 549.
- [5] RUPRICHOVÁ L. – DRAČKOVÁ M. – BORKOVCOVÁ I. et al. 2010. Determination whey proteins with allergenic potential in cow's and sheep's milk. In Text book of the international scientific conference. Nitra : SPU, p. 199 – 202.
- [6] STEFANIAK, T. – CHRZANOWSKA, J. – NIKOLAJCZUK, M. et al. 2000. Immunologiczne i elektroforezyczne wykrywanie zafalszowań mleka koziego i jego produktów mlekiem krowim. In *Medycyna Wet*, 56, 2000, 8, p. 505 – 509.
- [7] XUE Hai-yan – HU Wei-wei – SONG Hong-xin et al. 2010. Indirect ELISA for Detection and Quantification of Bovine Milk in Goat Milk. In *J. Food Sci. Tech.*, 31, 2010, 24, p. 370 – 373.
- [8] ZELENÁKOVÁ, L. – GOLIAN, J. – ZAJÁC, P. 2008. Application of ELISA tests for the detection of goat milk in shep milk. In *Milchwissenschaft*, 63, 2008, 2, p. 137 – 141.
- [9] ZELENÁKOVÁ, L. – ŠESTÁK, M. – ŽIDEK, R. 2009. Monitoring falšovania ovčieho mlieka a výrobkov z neho na spoločnom európskom trhu s potravinami. In *Potravinárstvo*, 3, 2009, 2, s. 69 – 73.

Adresa autorov:

Ing. Lucia Zelenáková, PhD.
Katedra hygieny a bezpečnosti potravín
Fakulta biotechnológií a potravinárstva
Slovenská poľnohospodárska univerzita
Trieda A. Hlinku 2, 949 76 Nitra
E-mail: lucia.zelenakova@uniag.sk

Ing. et Bc. Ladislav Mura, Ph.D.
Ústav ekonomie
Moravská vysoká škola Olomouc
Jeremenkova 42
772 00 Olomouc, Česká republika
E-mail: ladislav.mura@mvsso.cz

Obsah

Andrejiová, M., Kimáková, Z.	Plošný prieskum kontaminácie pôdy na Slovensku	2
Andrejiová, M., Kimáková, Z.	Regulačné diagramy a spôsobilosť výrobného procesu pomocou programu R	8
Balajty, M., Boďa, M.	Aplikácia metód viackriteriálneho hodnotenia pri skúmaní vplyvu ekonomicko-právnych aspektov na zakladanie nadnárodných obchodných spoločností v Európskej únii	14
Bízová, L.	Vztah spotřeby a objemu dlouhodobých úvěrů domácnostem v ČR	19
Boďa, M.	Inferencia v zovšeobecnenom gaussovskom lineárnom regresnom modeli v malých výberoch za predpokladu multiplikatívnej heteroskedasticity: pokračovanie štúdie Monte Carlo	25
Boďa, M., Roháčová, V.	K validite výsledkov obalovej analýzy dát pri meraní technickej efektívnosti	31
Brodňani, J.	Prognóza a verifikácia výkonov v mužských atletických disciplínach do OH v Londýne	37
Buleca, J., Šoltés, M., Mura, L., Kozelová, D.	Hodnotenie asociácie DNA markerov k úžitkovým vlastnostiam zvierat pomocou neurónových sietí	42
Fándlyová, S., Rybanský, L	Názory študentov na vyučovanie matematiky a prírodovedných predmetov	47
Hampel, D., Hanzelková, Š., Bubeničková, A.	Souvislost instrumentálních charakteristik sýrů zrajících pod mazem s délkou zrání	54
Chajdiak, J., Věntusová, K.	Vývoj tržieb v potravinárskom priemysle SR v období 2000 až 2010	59
Ivor, J., Stehlíková, B., Tirpáková, A.	Podobnosť krajín EÚ z hľadiska vnímania korupcie	65
Ižaričková, G., Olekšáková, D.	Porovnanie miery nezamestnanosti v Košickom regióne a v Slovenskej republike	70
Jenis, J., Jenis, J.	Simulácia systému hromadnej obsluhy v programe R	76
Kóšová, M.	Štatistická gramotnosť v súvislosti s polopravdami a zavádzajúcimi vyjadreniami v štatistike	80
Kurucová, Z., Stehlíková, B., Tirpáková, A.	Použitie Kruskalovho-Wallisovho testu pri overovaní účinnosti kombinovaného spôsobu vyučovania anglického jazyka	85
Laník, P.	Semi-parametrický prístup k odhadovaniu v autoregresnom modeli	91
Lašová, L.	Výhody grafickej knižnice lattice v programe R	97
Macháčová, E., Luha, J.	Názory študentov LF UK a FaF UK na fajčenie	103
Markechová, D.	Štatistické vyhodnotenie výsledkov výskumu v hudbe Friedmanovým testom	109
Mura, L.	Využitie softvéru Mathematica pri názornosti kriviek	114
Myšková, K.	Jednorozměrná kalibrace s kvadratickou kalibrační funkcí	120

Obtulovič, P., Poláková, Z.	Zmeny vo vývoji evidovanej miery nezamestnanosti v Nitrianskom kraji	124
Olekšáková, D., Ižariková, G.	Vplyv doby mletia a teploty žihania na hodnotu koercivity kompaktovaných vzoriek zisťovaný pomocou štatistických metód	129
Országhová, D., Gregáňová, R.	Analýza výsledkov skúšok z matematických povinných a voliteľných predmetov	133
Pánik, M.	Prognóza vývoja priemerných cien bytov v Slovenskej republike	138
Pobočíková, I. Wiszt, R.	Konfidenčný interval pre rozdiel parametrov dvoch nezávislých binomických rozdelení s aplikáciou v klinických pokusoch	144
Poláková, Z. Obtulovič, P.	Analýza a komparácia demografického vývoja rómskej populácie vo vybraných regiónoch SR	150
Potisková, L.	Oblúkový efekt korešpondenčnej analýzy	155
Radzo, E., Repiská, V., Luha, J.	Karcinóm endometria (CaE), analýza závislostí	161
Rumanová, L., Pavlovičová, G., Švecová, V.	Využitie štatistického softvéru C.H.I.C pri vyhodnení didaktického experimentu	166
Stehlíková, B., Markechová, D.	Štatistika - vstupná brána poznania	171
Šácha, J.	Kvazinormy s obecnými podmínkami	174
Tirpák, J.	Analýza výkonov Puncového úradu Slovenskej republiky v rokoch 1970 – 2010 použitím štatistických metód	180
Trnka, A., Kovářová, M., Dóci, I.	Výskyt sexuálnych porúch v SR – päťročná analýza pomocou doloženia dát	185
Uhrinová, E.	Porovnanie štatistickej gramotnosti krajín sveta	193
Vrábelová, M., Urbaníková, M.	Pravdepodobnostné rozdelenie likvidného rizika	199
Zeleňáková, L., Matejková, E., Lopašovský, E., Kozelová, D.	Hodnotenie mikrobiologickej kvality vína využitím štatistických metód	205
Zeleňáková, L., Mura, L.	Hodnotenie kvality imunoenzymatických testov pri detekcii vzájomného falšovania kravského a ovčieho mlieka	211

Content

Andrejiová, M., Kimáková, Z.	Area soil contamination survey in Slovakia	2
Andrejiová, M., Kimáková, Z.	Quality control charts and process capability in program R	8
Balajty, M., Boďa, M.	An application of multicriterial assessment methods in the investigation of the influence of economic-legislative factors upon the founding of transnational corporations in the European Union	14
Bízová, L.	The relationship of consumption and the volume of long-term loans to households in the Czech Republic	19
Boďa, M.	Small-sample inference in the general Gaussian linear regression model under the assumption of multiplicative heteroskedasticity: continuation of a Monte Carlo study	25
Boďa, M., Roháčová, V.	To the validity of DEA results in technical efficiency measurement	31
Broďáni, J.	Prognosis and verification performance in male athletic disciplines to the Olympic Games in London	37
Buleca, J., Šoltés, M., Mura, L., Kozelová, D.	Evaluation of DNA markers association in quality traits using neural networks	42
Fándlyová, S., Rybanský, Ľ	Academic opinion of mathematical and science education	47
Hampel, D., Hanzelková, Š., Bubeničková, A.	Relation of instrumental characteristics of acid smeared cheeses to ripening time	54
Chajdiak, J., Věntusová, K.	Food Industry turnover development in the period from 2000 to 2010	59
Ivor, J., Stehlíková, B., Tirpáková, A.	Similarity of EU countries in terms of corruption perception	65
Ižaríková, G., Olekšáková, D.	The comparison of the unemployment rate in Kosice region and Slovak republic	70
Jeníš, J., Jeníš, J.	The Simulation of a Bulk Service System in Program R	76
Kóšová, M.	Statistical literacy in connection with half-truths and confusing statements in statistics	80
Kurucová, Z., Stehlíková, B., Tirpáková, A.	The Application of the Kruskal-Wallis' Test in the Testing of Efficiency of Combined Methods of Teaching of Professional English	85
Laník, P.	Semi-parametric approach for fitting of autoregressive model	91
Lašová, L.	The advantages of the package lattice in software R	97
Macháčková, E., Luha, J.	Opinion of students LF UK and FaF UK on smoking	103
Markechová, D.	Statistical evaluation of results of the research in music by Friedman's test	109
Mura, L.	Using of Mathematica software by visualization of curve	114

Myšková, K.	One-dimensional calibration with quadratic calibration function	120
Obtulovič, P., Poláková, Z.	Changes in the development registered unemployment rate in the Nitra Region	124
Olekšáková, D., Ižaričková, G.	The influence of the time of milling and the temperature of annealing to the value of the coercivity of compacted samples investigated by statistical methods	129
Országhová, D., Gregáňová, R.	The analysis of exams outcomes of the mathematical compulsory and optional subjects	133
Pánik, M.	Prognosis of the average apartment prices in Slovakia	138
Pobočíková, I., Wiszt, R.	The confidence interval for the difference of two independent binomial proportions with application in the clinical trials	144
Poláková, Z., Obtulovič, P.	Analysis and comparison of the demographic development of the Roma population in selected regions of Slovakia	150
Potisková, L.	Arch effect in Correspondence Analysis	155
Radzo, E., Repiská, V., Luha, J.	Carconoma endometrium (CaE), analysis of dependencies	161
Rumanová, L., Pavlovičová, G., Švecová, V.	Apply of statistical software C.H. I.C in evaluation of didactical experiment	166
Stehlíková, B., Markechová, D.	Statistics - a gateway of knowledge	171
Šácha, J.	Quasi-norms with general constraints	174
Tirpák, J.	Performances of the Assay Office of the Slovak republic in the years 1970 – 2010	180
Trnka, A., Kovářová, M., Dóci, I.	Occurrence of sexual disorders in Slovakia - a five-year analysis with Data Mining	185
Uhrinová, E.	Comparison of Statistical Literacy in countries around the World	193
Vrábelová, M., Urbaníková, M.	The Probability Distribution of The Liquidity Risk	199
Zeleňáková, L., Matejková, E., Lopašovský, Ľ., Kozelová, D.	Assessment of microbiological quality wines using statistical methods	205
Zeleňáková, L., Mura, L.	Quality evaluation of immuno enzymatic tests for the detection of mutual adulteration cow and sheep milk	211