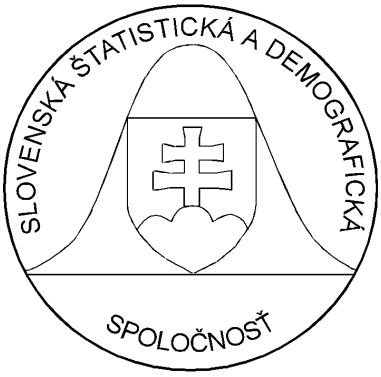
1/2009

FORUM STATISTICUM SLOVACUM



ISSN 1336-7420
91
9 771336 742001

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

Naše najbližšie akcie:

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

Pohľady na ekonomiku Slovenska 2009
7. 4. 2009, Bratislava

EKOMSTAT 2009, 23. škola štatistiky
tematické zameranie: *Štatistické metódy vo vedecko-výskumnej, odbornej a hospodárskej praxi.*
31. 5. 2009 – 5. 6. 2009, Trenčianske Teplice

Aplikácie metód na podporu rozhodovania vo vedeckej, technickej a spoločenskej praxi
jún 2009, STU Bratislava

Slovensko-česká konferencia PRASTAN
jún 2009 Kočovce

12. SLOVENSKÁ DEMOGRAFICKÁ KONFERENCIA
tematické zameranie: Demografická budúcnosť Slovenska
23. – 25. 9. 2009, Bojnice

FernStat 2009
VI. medzinárodná konferencia aplikovanej štatistiky
(**F**inancie, **E**konomika, **R**iadenie, **N**ázory)
tematické zameranie: *Aplikovaná, demografická, matematická štatistika, štatistické riadenie kvality*.
október 2009, hotel Lesák, Tajov pri Banskej Bystrici

18. Medzinárodný seminár VÝPOČTOVÁ ŠTATISTIKA,
3. – 4. 12. 2009, Bratislava, Infostat

Prehliadka prác mladých štatistikov a demografov
3. 12. 2009, Bratislava, Infostat

Regiónálne akcie
priebežne

ÚVOD

Vážené kolegyne, vážení kolegovia,

prvé číslo piateho ročníka vedeckého časopisu Slovenskej štatistickej a demografickej spoločnosti je zostavené z príspevkov, ktoré sú obsahovo orientované v súlade s tematikou konferencie Nitrianske štatistické dni. Táto konferencia sa uskutočnila v Nitre, v dňoch 5. a 6. februára 2009. Konferenciu organizovala Slovenská štatistická a demografická spoločnosť v spolupráci s Fakultou prírodných vied UKF v Nitre.

Akciu z poverenia Výboru SŠDS, zorganizoval Organizačný a programový výbor: doc. RNDr. Anna Tirpáková, CSc. – predseda, PaedDr. Janka Melušová, PhD. – tajomník, prof. RNDr. Ondrej Šedivý, CSc., doc. RNDr. Bohdan Linda, CSc., doc. PaedDr. Jana Kubanová, CSc., Ing. Iveta Stankovičová PhD., doc. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc., doc. RNDr. Dagmar Markechová, CSc., RNDr. Kitti Vidermanová, PhD.

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

Recenziu príspevkov zabezpečili: doc. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc., doc. RNDr. Bohdan Linda, CSc., doc. PaedDr. Jana Kubanová, CSc., doc. RNDr. Dagmar Markechová, CSc., PaedDr. Janka Melušová, PhD., doc. RNDr. Anna Tirpáková, CSc., PaedDr. Lucia Záhumenská.

Výbor SŠDS

Využitie mnohonásobnej korelačnej a regresnej analýzy časových radov vo vrcholovom športe

Utilization of Multiple Correlation and Regresion Analysis of the Time Series in Top Sport

Jaroslav Broďáni

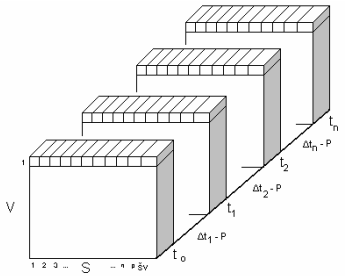
Abstract: The paper deals with the problem of the use of multiple correlative and regressive analysis of temporary sequences in top level sport. Attention is focused on the area of intraindividual research, or on solving the problem of the efficiency of training load on a sport performance.

Key words: Time series, correlation and regresion analysis, top sport

Kľúčové slová: Časové rady, korelačná a regresná analýza, vrcholový šport

1. Úvod

Objektivizácia účinku tréningových prostriedkov je často dôležitým zámerom vedeckého skúmania v oblasti vrcholového športu. V empirickom výskume sa uplatňujú dva prístupy založené na interindividuálnom a intraindividuálnom postupe (obr. 1). Prvý z nich je zameraný na sledovanie viacpočetných objektov sledovania a ich vnútornej variabilite javov a procesov. Jedinečnosť človeka vo všetkých stránkach jeho existencie si však vyžaduje osobitný prístup skúmania. Sledovanie jednotlivcov prináša priame vedecké poznatky o účinkoch intervencie v určitých časových intervaloch, ktorými možno odhaliť všeobecné vzťahy aplikovateľné na širšiu skupinu s určitou mierou homogenity (Havlíček - Záhorec, 1999).



Obrázok 1: Empirický model intraindividuálnej výskumnej situácie so znalosťou podnetov v trojrozmernom priestore

Vzťah charakterizujúci výskumnú situáciu:

V_1: S, $t_0 \rightarrow$ ((P_1, ... P_n) Δt_1) $\rightarrow$ S, $t_1 \rightarrow$ ((P_1, ... P_n) Δt_2) $\rightarrow$ S, $t_2 \rightarrow$ ..... $\rightarrow$ S, $t_{60} \rightarrow$ ((P_1, ... P_{61}) Δt_{61}) $\rightarrow$ S, $t_{61} \rightarrow$ ((P_1, ... P_n) Δt_n) $\rightarrow$ S, t_n

Intraindividuálny výskum (obr. 1) stanovujú podmienky, pri ktorých výskum prebieha čo do rozsahu výberu probantov „V“ a ich stavov „S“, času sledovania „Δt“ a pôsobiacich

podnetov „P. Táto výskumná situácia nám umožňuje: 1. Zistiť dynamiku zmien úrovne tréningového zaťaženia a zmien športového výkonu. a 2. Overiť účinnosť tréningového zaťaženia na zmenách športového výkonu.

Jeho komplexné využitie v empirickom výskume umožňuje prispieť k objektivizácii vzťahu medzi tréningovým zaťažením a zvolenými kritériami výkonnosti. Umožňuje poukázať na jedinečnosť športovca, jeho individuality v prostredí vrcholovej športovej prípravy. S využitím párovej a mnohonásobnej korelačnej a regresnej analýzy časových radov v intraindividuálnom výskume sú známe práce autorov, ako Belás, 2006; Broďáni, 2008; Čillík, 1993; Glesk, 1987; Havlíček - Záhorec, 1983; Horička, 2006; Hutňan, 1987; Kolčiter, 1993; Krška – Koštial, 2001; Laczo, 1987; Macejková, 1999; Matoušek, 1988; Ryšavý, 1995; Vavák, 1999; Vindušková, 1987; Záhorec, 1989, atď.

Z hľadiska odhaľovania faktorov podmieňujúcich úroveň športovej výkonnosti, resp. hodnotenia účinnosti tréningového zaťaženia, jeho prostriedkov poskytuje rozšírené možnosti objektivizácie poznatkov využitie párovej a mnohonásobnej korelačnej a regresnej analýzy časových radov (Havlíček - Záhorec, 1983). Využitie tohto postupu v empirickom výskume umožňuje prispieť k objektivizácii vzťahu medzi tréningovým zaťažením a zvolenými kritériami výkonnosti. Umožňuje poukázať na jedinečnosť športovca, jeho individuality v prostredí vrcholovej športovej prípravy.

V oblasti skúmania štruktúry športového výkonu, faktorov, ktoré determinujú jeho úroveň a vzťahov medzi nimi je aplikáciou intraindividuálneho prístupu z hľadiska časových parametrov charakteristická dominancia synchrónnosti snímania údajov nezávislých a závislých premenných (Bakytová a kol., 1979). Asynchrónny pohľad umožňuje priniesť poznatky o niektorých čiastkových smeroch, kedy ukazovatele nezávislých premenných pri ovplyvnení športového výkonu presahujú pásmo determinačného optima. Aplikácia párovej, resp. mnohonásobnej korelačnej a regresnej analýzy časových radov je potrebné dodržať niektoré parametre ako dostatočná variabilita primárnych údajov, optimálny počet časových intervalov, ekvidištančnosť primárnych údajov, čo najmenší počet doplňaných údajov, výber prijateľnej autokorelácie, rozhodnutie o eliminácii trendovej zložky.

Problematickým momentom je v tomto prípade rozhodnutie o použití alebo nepoužití očistených časových radov od trendu. Rozhodnutie sa zakladá na dvoch prístupoch. Prvý vychádza z predpokladu, že regresia časových radov je možná, pokiaľ medzi nimi je kauzálny vzťah, pričom rešpektujeme originalitu údajov. Druhý názor tvrdí, že regresia časových radov v každom prípade začína očistením časových radov od trendu, čím eliminujeme poruchy regresie vyplývajúce z trendom zaťažených radov (Záhorec, 2001). Oba prístupy majú svoje opodstatnenie. Prvý prístup sa osvedčil v prípade zisťovania extrapolovaných prognostických hodnôt. Druhý prístup má opodstatnenie pri zisťovaní všeobecnej pravidelnosti, zákonitosti vzťahu medzi kriteriálnymi (vysvetľovanými) a vysvetľujúcimi premennými. V prípade použitia korelačnej a regresnej analýzy časových radov je nutné akceptovať druhý prípad (Bakytová a kol., 1979; Cipra, 1986; Chajdiak, 2002).

2. Analýza výskumu účinnosti tréningového zaťaženia

V oblasti výskumu účinnosti tréningového zaťaženia, jeho prostriedkov je aplikácia intraindividuálneho prístupu z hľadiska časových parametrov charakterizovaná dominanciou asynchrónneho usporiadania údajov nezávislých a závislých premenných. Preukázateľný efekt, účinnosť tréningového zaťaženia ako celku alebo jednotlivých prostriedkov, vo vzťahu k rozdielnym kriteriálnym ukazovateľom je veľmi zložitý. Pedagogický pohľad považuje za účinné tréningové zaťaženie také, ktoré z hľadiska objemu, intenzity a zložitosti parametrov a dávkovania vyvoláva účinok (efekt) väčší ako iné postupy, vytvára bázu pre dlhodobý výkonnostný rast, negatívne neovplyvňuje zdravie športovca (Koštiaľ, 1984). Z pohľadu užšej

špecifikácie je možné za účinnosť tréningového zaťaženia považovať vzťah medzi objemom, intenzitou a zložitosťou tréningového zaťaženia, jeho jednotlivých prostriedkov a úrovňou a zmenami kriteriálnych faktorov odrážajúcich sa v úrovni a zmenách športového výkonu (Záhorec, 2001).

Prevažná väčšina používaných prostriedkov tréningového zaťaženia je z hľadiska popisu pohybovej činnosti a z hľadiska objemu a intenzity charakterizovaná relatívne presne (tréningové denníky). Kvantita a kvalita registrácií informácií o aplikácii jednotlivých prostriedkov tréningového zaťaženia v tréningovom procese podmieňuje možnosť čo najobjektívnejšie posúdiť ich účinnosť na požadovanej úrovni. Z tohto dôvodu sú dôležité kritériá - parametre pre ich registráciu: presné definovanie prostriedkov tréningového zaťaženia, spôsob registrácie a ich aplikácie z hľadiska kvality a kvantity a stanovenie kritéria účinnosti.

Tabuľka 1 Príklad korelácie podnetov (tréningových ukazovateľov) k zmenám športovému výkonu u chodca - olympionika na 20 km z časového obdobia 2001-2006, (Broďáni, 2008).

		Časový odstup od športového výkonu (4-týždňové mezocykly)								
Tréningové ukazovatele / posun		0	1	2	3	4	5	6	7	8
115	Dni zaťaženia (n)		*	**					*	
116	Tréningové jednotky (n)		**	**		***			*	
119	Regenerácia síl [min]		**			****				*
101	Chôdza pod 3:40 min.km^{-1} [km]		**	***	****					
102	Chôdza 3:41 - 4:05 min.km^{-1} [km]		**		*					
103	Chôdza 4:06 - 4:20 min.km^{-1} [km]	**								
104	Chôdza 4:21 - 4:40 min.km^{-1} [km]	*								
105	Chôdza 4:41 - 5:00 min.km^{-1} [km]		*	***		***	*	***	*	*
106	Chôdza 5:01 - 5:20 min.km^{-1} [km]		*				*		**	
107	Chôdza 5:21 - 5:40 min.km^{-1} [km]								*	
108	Chôdza 5:41 - 6:00 min.km^{-1} [km]			**		****		**		
109	Chôdza 6:00 a viac min.km^{-1} [km]			**	**	**	***			*
110	Súčet chôdza [km]		**	**		***			**	*
111	Súčet beh [km]		***	*		***		**		
113	Celkový objem [km]		**	*		**			**	*
114	Doplnky VTP [min]	****		**						

Legenda:

Kladná hladina významnosti	***20% ****	***10 %*****	***5%******	***1 % *******
Záporná hladina významnosti	***20% ****	***10 %*****	***5%******	***1 % *******

Stanovenie účinnosti prostriedkov tréningového zaťaženia sa deje v súlade so stanovením kritéria. Základným kritériom sú zmeny v úrovni športového výkonu, resp. faktorov, ktoré športový výkon podmieňujú. Prístupy, metódy a formy, ktoré používame pri získavaní informácií o zmenách v hodnotách kriteriálnych ukazovateľov musia tiež spĺňať požiadavky na možnosti spracovania a vyhodnotenia účinnosti prostriedkov (validita, objektivita, spoľahlivosť, citlivosť, jednoduchosť, časovo-materiálna úspornosť, atď.). Vhodnosť vecného posúdenia priameho alebo nepriameho vzťahu medzi zmenami kritéria a zmenami ukazovateľov je jednou z ďalších požiadaviek.

Pre spracovanie a vyhodnocovanie údajov pomocou párovej a mnohonásobnej korelačnej a regresnej analýzy časových radov platia rovnaké podmienky ako pri tvorbe štruktúry športového výkonu. Dominantnú úlohu majú pritom originálne údaje (primárne údaje). Posudzovanie vplyvu autokorelácií vo vzťahu k zdroju variability môže byť miernejšia

(primárne údaje, reziduálne zložky časových radov). Výber aproximačnej (vyrovnávajúcej funkcie vo vzťahu k vecne-logickému hľadisku konkrétnej výskumnej situácie. Využitie viacnásobného asynchrónneho usporiadania nezávisle a závisle premenných pri hľadaní optimálneho časového odstupu medzi pôsobením príčiny (tréningové zaťaženie) a dostavením sa relatívne najvyššieho účinku na kriteriálny ukazovateľ.

Korelácie primárnych údajov (originálnych) a reziduálnych zložiek časových radov (eliminácia trendovej zložky) však často presahujú úroveň štatistickej významnosti aj na menej prísnych hladinách (10-20%). Miera ovplyvnenia interpretácie výsledkov v kauzálnom smere sa stáva viac-menej nekontrolovateľnou. Evidentný sa preukazuje vplyv úprav pomocou fitovania primárnych údajov kubickými spline funkciami. Zdôrazňuje sa tu výber funkcie na elimináciu trendovej zložky.

Oblasť využitia analýzy časových radov v športe nám umožňuje tri oblasti skúmania:

Oblasť hodnotenia účinnosti tréningového zaťaženia a jednotlivých prostriedkov z hľadiska ich kumulatívneho vplyvu. Samotné hodnotenie si pritom vyžaduje špecifický prístup k nevyhnutnej eliminácii trendovej zložky rozšírený o hľadisko výlučne neklesajúceho charakteru údajov v ukazovateľoch nezávisle premenných nasledujúcich za sebou.

Oblasť účinnosti tréningového zaťaženia v rámci športovej prípravy mládeže a vrcholového športu pri rešpektovaní odlišnej úrovne, priebehu a variability zmien hodnôt závislých a nezávislých ukazovateľov. Príčina v tomto prípade môže byť odlišná a taktiež nie výsledok. Nevyhnutný je tu výber správnej vyrovnávajúcej funkcie, ktorá eliminuje trend.
Oblasť predchádzajúceho vplyvu úrovne športového výkonu ako nasledujúcu determinantu toho istého kritéria. V matematicko-štatistickom ponímaní sa jedná o využitie autokorelácií prvého a vyšších rádov.

3. Záver

V konečnom dôsledku treba upozorniť na rozhodujúcu úlohu rešpektovania nielen matematicko-štatistických metód ale predovšetkým vecne-logickej analýzy konkrétnej výskumnej situácie, jej konštrukcie vo vzťahu možnostiam získavania dostatočne informatívnych, objektívnych a spoľahlivých výsledkov. Rešpektovanie poznatkov z hľadiska štruktúry časovo súbežných a nesúbežných tréningových prostriedkov a dynamiky zmien objemu zaťaženia intraindividuálnej adaptácie nám umožňuje kvalitné plánovanie zaťaženia a jeho intenzifikáciu v období ladenia vrcholnej športovej formy.

4. Literatúra

[1] BAKYTOVÁ, H. - UGRON, M. - KONTEŠKOVÁ, O. 1979. Základy štatistiky. Bratislava : Alfa, 1979.

[2] BELÁS, M. 2006. Intraindividuálne hodnotenie účinnosti tréningových prostriedkov v horskej cyklistike. (KDP). Bratislava: FTVŠ UK, 2006, 91 s.

[3] BROĎÁNI, J. 2008. Účinnosť tréningového zaťaženia na športový výkon v dlhodobej športovej príprave u chodca na 20 km: habilitačná práca. – Prešov, [s.n.], 2008, 109 s.

[4] CIPRA,T. 1986. Analýza časových řad s aplikacemi v ekonomii. Praha : Alfa, 1986.

[5] ČILLÍK, I. 1993. Efektivita tréningového zaťaženia v skoku do diaľky. [KDP]. Bratislava : FTVŠ UK.

[6] GLESK, P. 1987. Retrospektívna analýza účinnosti tréningového zaťaženia. In Sborník vědecké rady ÚV ČSTV, 1987, č. 18.

[7] HAVLÍČEK, I. - ZÁHOREC, J. 1983. Hodnotenie vzťahu medzi tréningovou záťažou a športovým výkonom korelačnou analýzou časových radov. In Teor. a Prax. Těl. Vých., 31, 1983, č. 1, str. 9-16.

[8] HAVLÍČEK, I. - ZÁHOREC, J. 1999. Uplatnenie intraindividuálnej metodológie v empirickom výskume. In 60 rokov prípravy telovýchovných pedagógov na UK v Bratislave. Bratislava : FTVŠ UK s. 28-31.

[9] HORIČKA, P. 2006. Vzťah vybraných tréningových prostriedkov a motorickej výkonnosti v basketbale. In Optimální působení tělesné zátěže a výživy. Hradec Králové : PF FIM, 2006, s. 236-242. ISBN 80-7041-104-X.

[10] HUTŇAN, D. 1987. Využitie korelačnej analýzy časových radov pri hodnotení dynamiky objemovej prípravy vrcholových žrďkárov. Teor. a praxe těl. výchovy, 35, 1987, č.1.

[11] CHAJDIAK, J. 2002. Analýza časových radov. In: Štatistika v exceli. Statis Bratislava 2002, 159 s. ISBN 80-85659-27-1.

[12] KOSŤIAĽ, J. 1984. Účinnosť tréningového zaťaženia na pohybové schopnosti a výkonnosť mládeže v atletike. [KDP]. Bratislava : FTVS UK, 1984.

[13] KOLČITER, J. 1993. Intraindividuálny výkonnostný efekt tréningového zaťaženia skokanov do diaľky. [KDP] Bratislava: FTVS UK, 1993.

[14] KRŠKA, P. - KOŠTIAL, J. 2001. Intraindividuálna účinnosť tréningového zaťaženia na zmeny kondičnej pripravenosti a športovej výkonnosti skokanky o žrdi. In: Optimalizácia zaťaženia v telesnej a športovej výchove. Bratislava : STU, 2001, s. 76 - 81. ISBN 80-227-1541-7

[15] LACZO, E. 1987. Zvyšovanie športového výkonu v závislosti od rozvoja pohybových schopností a tréningového zaťaženia v prekážkovom šprinte. [KDP]. Bratislava : FTVŠ UK, 1987.

[16] MACEJKOVÁ, Y. 1999. Účinnosť tréningové zaťaženia na športový výkon vrcholových plavkýň. [Habilitačná práca]. Bratislava: FTVŠ UK, 1999.

[17] MATOUŠEK, R. 1988. Účinnosť tréningového zaťaženia v prekážkových behoch. [KDP]. Bratislava: FTVŠ UK, 1988.

[18] RYŠAVÝ, B. 1995. Adaptačný efekt tréningového zatížení sprinterú. [KDP]. Praha : FTVS UK, 1995.

[19] VAVÁK, M. 1999. Intraindividuálna závislosť športového výkonu na tréningovom zaťažení v atletickej chôdzi. [KDP]. FTVŠ UK, Bratislava 1999, 130 str.

[20] VINDUŠKOVÁ, J. 1987. Rozbor tréningového zatížení na základe diachronné intrainduividuální variability zkoumaných ukazovatelú. [KDP]. Praha : FTVS KU, 1987.

[21] ZÁHOREC, J. 1989. Intraindividuálne hodnotenie účinnosti tréningového zaťaženia: KDP. Bratislava, [s.n.], 1989, 145 str.

[22] ZÁHOREC, J. 2001. Intraindividuálna analýza štruktúry športového výkonu a účinnosti tréningového zaťaženia. [Habilitačná práca], Bratislava: FTVŠ UK, 2001, 120 s.

Adresa autora:
Jaroslav Broďáni, Doc., PaedDr., PhD.
Tr. A. Hlinku 1
974 01 Nitra
jbrodani@ukf.sk

Štatistické riadenie kvality - Stratifikované náhodné vyberanie
Statistical Quality Control - Stratified Random Sampling

Hrnčiarová Ľubica

Abstract: Sampling is an integral part of statistical quality control. The author's attention in this paper will be focused on stratified random sampling**.**

Key words: Sample survey, sampling design, finite population sample, a sample, the point and interval estimate of mean value and grand total for population sample, determination of sample size.

Kľúčové slová: Výberové skúmanie, plán výberového skúmania, konečný základný súbor, výberový súbor, bodový a intervalový odhad strednej hodnoty a úhrnu v základnom súbore, stanovenie rozsahu výberu.

1. Úvod

Vznik v súčasnosti používaných metód vyberania možno datovať na začiatok 20-teho storočia. Ich používanie sa stalo bežnou záležitosťou. Využívajú sa pri väčšine rozhodnutí v oblasti vládnej politiky, marketingu (vrátane obchodovania) a výroby. Vyberanie je integrálnou súčasťou aj štatistického riadenia kvality (SQC). Sú známe tri rozličné prístupy k výberovému skúmaniu, a to prístup založený na pláne výberového skúmania (design-based approach), prístup založený na modeli (model-based approach) a prístup s asistenciou modelu (model-assisted approach). V príspevku, z plánov výberového skúmania sa budem zaoberať stratifikovaným náhodným vyberaním.

Primárnym cieľom vyberania je odhad parametrov základného súboru na základe informácií obsiahnutých vo výberovom súbore.

V ďalšom texte príspevku je popísaný postup stratifikovaného náhodného vyberania, bodový a intervalový odhad strednej hodnoty, úhrnu v základnom súbore a určenie minimálneho rozsahu výberu. V závere príspevku je uvedený príklad odhadu strednej hodnoty a úhrnu v konečnom základnom súbore.

2. Stratifikované náhodné vyberanie

V hospodárskej praxi nie je ničím výnimočným, že pri porovnávaní hodnôt skúmaného znaku medzi jednotlivými štátmi, regiónmi, podnikmi a pod. sa zistia rozdiely. Napríklad, zaujímala by nás životnosť určitého výrobku vyrábaného vo viacerých podnikoch. Životnosť výrobku môže byť rozdielna. Môže to byť spôsobené napríklad vstupnými surovinami, odbornou úrovňou pracovníkov, technickou úrovňou strojového parku. V takýchto prípadoch základný súbor je nehomogénny (heterogénny) a presnejšie odhady parametrov základného súboru získame, keď použijeme stratifikované náhodné vyberanie. Pri stratifikovanom vyberaní sa najprv základný súbor rozdelí na vzájomne sa vylučujúce a základný súbor celkom pokrývajúce podsúbory (stratá alebo vrstvy), ktoré sa vzhľadom na skúmaný znak považujú za viac homogénne ako celý základný súbor. Potom sa zrealizuje v každom strate náhodné vyberanie.

Niektoré z výhod stratifikovaného náhodného vyberania oproti jednoduchému náhodnému vyberaniu sú nasledovné:

1. poskytuje presnejšie odhady parametrov základného súboru ako by mohlo poskytovať jednoduché náhodné vyberanie s rovnako veľkým rozsahom výberu.
2. stratifikované vyberanie je vhodnejšie na aplikáciu, pretože môžeme dosiahnuť výsledky s nižšími nákladmi na vyberanie.

3. dostaneme jednoduché náhodné výbery pre každú podskupinu (stratum), ktoré sú homogónnejšie. Preto, tieto výbery môžeme využiť užitočnejšie na študovanie každej individuálnej podskupiny samostatne bez prípadných nákladov naviac.

Predtým ako sa pozrieme na to ako využiť informácie získané zo stratifikovaného náhodného vyberania na odhad parametrov základného súboru, stručne preberieme proces vytvorenia stratifikovaného náhodného výberu:

- Základný súbor rozsahu N jednotiek rozdelíme na vzájomne sa vylučujúce a základný súbor celkom pokrývajúce podsúbory (*stratá* alebo *vrstvy*) rozsahu N_1, N_2, ..., N_m jednotiek , pričom $N = N_1 + N_2 + ... + N_m$.
- Vyberieme z každého strata jednoduchý náhodný výber rozsahu n_1, n_2,..., n_m tak, aby $n = n_1 + n_2 + ... + n_m$ bol celkový rozsah výberu.
- Aby sme mohli použiť stratifikáciu, rozsah jednotlivých strat N_1, N_2, ..., N_k musí byť známi.

2.1 Odhad strednej hodnoty a úhrnu v konečnom základnom súbore

Nech $X_{11}, X_{12}, ..., X_{1N_1}; X_{21}, X_{22}, ..., X_{2N_2}; X_{m1}, X_{22}, ..., X_{mN_m}$ sú výberové jednotky v 1, 2, ..., m- tom strate a nech $X_{11}, X_{12}, ..., X_{1n_1}; X_{21}, X_{22}, ..., X_{2n_2}; X_{m1}, X_{22}, ..., X_{mn_m}$ sú jednoduché náhodné výbery rozsahu n_1, n_2,..., n_m. Nech μ_K je stredná hodnota a τ úhrn v konečnom základnom súbore a μ_h je stredná hodnota, τ_h úhrn a σ_h^2 rozptyl v h – tom strate. Potom dostaneme

$$\tau = \sum_{h=1}^{m} \tau_h = \sum_{h=1}^{m} N_h \mu_h , \tag{1}$$

$$\mu_K = \frac{\tau}{N} = \frac{1}{N} \sum_{h=1}^{m} N_h \mu_h . \tag{2}$$

Nech $\overline{X}_h$ je výberový priemer jednoduchého náhodného výberu v h – tom strate. Výberový priemer jednoduchého náhodného výberu v h - tom strate je nevychýleným odhadom strednej hodnoty μ_h v h – tom strate. Nevychýleným odhadom strednej hodnoty μ_K konečného základného súboru výberový priemer stratifikovaného výberu $\overline{X}_{str}$:

$$\hat{\mu} = \overline{X}_{str} = \frac{1}{N} \sum_{h=1}^{m} N_h \overline{X}_h , \tag{3}$$

kde

$$\overline{X}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} X_{hi} . \tag{4}$$

Odhadom úhrnu τ v základnom súbore je

$$\hat{\tau} = \sum_{h=1}^{m} \hat{\tau}_h = \sum_{h=1}^{m} N_h \hat{\mu}_h = \sum_{h=1}^{m} N_h \overline{X}_h = N \overline{X}_{str} . \tag{5}$$

Rozptyl $\overline{X}_{str}$ (odhadu strednej hodnoty μ_K v konečnom základnom súbore) je

$$D(\overline{X}_{str}) = \frac{1}{N^2} \sum_{h=1}^{m} N_h^2 D(\overline{X}_h) = \frac{1}{N^2} \sum_{h=1}^{m} N_h^2 \left(\frac{N_h - n_h}{N_h - 1} \right) \frac{\sigma_h^2}{n_h} . \tag{6}$$

Odhadom rozptylu $\overline{X}_{str}$ je

$$\hat{D}(\overline{X}_{str}) = \frac{1}{N^2} \sum_{h=1}^{m} N_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{S_h^2}{n_h} , \tag{7}$$

kde $$S_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left(X_{hi} - \overline{X}_h\right)^2 . \quad (8)$$

Rozptyl odhadu $\hat{\tau}$ je

$$D\left(\hat{\tau}_{str}\right) = N^2 D\left(\overline{X}_{str}\right) = \sum_{h=1}^{m} N_h^2 D\left(\overline{X}_h\right) = \sum_{h=1}^{m} N_h^2 \left(\frac{N_h - n_h}{N_h - 1}\right)\frac{\sigma_h^2}{n_h} . \quad (9)$$

Odhad rozptylu odhadu $\hat{\tau}$ je

$$\hat{D}\left(\hat{\tau}_{str}\right) = N^2 \hat{D}\left(\overline{X}_{str}\right) = \sum_{h=1}^{m} N_h^2 \left(1 - \frac{n_h}{N_h}\right)\frac{\sigma_h^2}{n_h} . \quad (10)$$

Ak rozdelenie v základnom súbore je normálne alebo rozsah výberu $n \geq 30$, potom prípustná chyba odhadu Δ strednej hodnoty μ_K konečného základného súboru s pravdepodobnosťou $(1-\alpha)$ je

$$\pm u_{1-\alpha/2}\frac{1}{N}\sqrt{\sum_{h=1}^{m} N_h^2 \left(\frac{N_h - n_h}{N_h - 1}\right)\frac{\sigma_h^2}{n_h}} . \quad (11)$$

Keď nepoznáme rozptyl σ_h^2 v jednotlivých stratách, nahradíme ho odhadom S_h^2. Potom prípustná chyba odhadu Δ strednej hodnoty μ_K konečného základného súboru s pravdepodobnosťou $(1-\alpha)$ je

$$\pm u_{1-\alpha/2}\frac{1}{N}\sqrt{\sum_{h=1}^{m} N_h^2 \left(1 - \frac{n_h}{N_h}\right)\frac{S_h^2}{n_h}} . \quad (12)$$

Prípustná chyba odhadu Δ úhrnu τ základného súboru s pravdepodobnosťou $(1-\alpha)$ je

$$\pm u_{1-\alpha/2}\sqrt{\sum_{h=1}^{m} N_h^2 \left(\frac{N_h - n_h}{N_h - 1}\right)\frac{\sigma_h^2}{n_h}} . \quad (13)$$

Keď nepoznáme rozptyl σ_h^2 v jednotlivých stratách, nahradíme ho odhadom S_h^2 a prípustná chyba odhadu Δ úhrnu τ v základnom súbore s pravdepodobnosťou $(1-\alpha)$ je

$$\pm u_{1-\alpha/2}\sqrt{\sum_{h=1}^{m} N_h^2 \left(1 - \frac{n_h}{N_h}\right)\frac{S_h^2}{n_h}} . \quad (14)$$

2.2 Intervaly spoľahlivosti pre strednú hodnotu a úhrn v konečnom základnom súbore

$100\cdot(1-\alpha)\,\%$ interval spoľahlivosti pre strednú hodnotu μ_K v konečnom základnom súbore, keď nepoznáme rozptyl σ_h^2 v jednotlivých stratách je

$$\overline{x_{str}} = \pm u_{1-\alpha/2}\frac{1}{N}\sqrt{\sum_{h=1}^{m} N_h^2 \left(1 - \frac{n_h}{N_h}\right)\frac{S_h^2}{n_h}} . \quad (15)$$

$100\cdot(1-\alpha)\,\%$ interval spoľahlivosti pre úhrn τ v konečnom základnom súbore, keď nepoznáme rozptyl σ_h^2 v jednotlivých stratách je

$$N\overline{X}_{str} \pm u_{1-\alpha/2}\sqrt{\sum_{h=1}^{m} N_h^2\left(1-\frac{n_h}{N_h}\right)\frac{S_h^2}{n_h}}. \qquad (16)$$

Poznámka

- Odhad $\overline{X}$ strednej hodnoty v základnom súbore z jednoduchého náhodného výberu sa zhoduje s odhadom $\overline{X}_{str}$ získaného zo stratifikovaného výberu, keď v každom strate $\frac{n_h}{N_h}$ sa rovná $\frac{n}{N}$.
- Ak rozsah výberu v každom strate *k* rozsahu tohto strata v základnom súbore je malý $\left(\frac{n_h}{N_h} < 5\%\right)$, potom $\frac{n_h}{N_h}$ možno vo vzorcoch vynechať.

Príklad 2.1 Predpokladajme, že firma vyrába súčiastky vo svojich troch výrobných prevádzkach, ktoré sa nenachádzajú na území jedného štátu a náklady na pracovníkov, suroviny a režijné náklady sú v jednotlivých prevádzkach veľmi rozdielne. Aby dosiahla cieľovú (stanovenú) hodnotu nákladovosti, zaujíma sa o odhad priemernej nákladovosti za sledované obdobie. Firma očakáva, že sa v jednotlivých výrobných prevádzkach za sledované obdobie vyrobí počet súčiastok $N_1 = 8900$, $N_2 = 10600$, $N_3 = 15500$. Očakávaný celkový počet vyrobených súčiastok za sledované obdobie za všetky prevádzky je $N = 8900 + 10600 + 15500 = 35000$ kusov. K dosiahnutiu stanoveného cieľa, zistili náklady na výrobu 12 náhodne vybratých súčiastok v prvej prevádzke, 12 náhodne vybratých súčiastok v druhej prevádzke a 16 náhodne vybratých súčiastok v tretej prevádzke (tab.2.1).

Tabuľka 2.1: Stratifikovaný náhodný výber

Stratum (prevádzka) h	Rozsah základného súboru N_h	Rozsah výberu n_h	Náklady na výrobu (v EUR) x_{hi}								
1	8900	12	4,73	4,88	5,19	4,49	5,18	5,31	5,53	4,74	4,61
			4,72	4,84	5,04						
2	10600	12	7,34	7,48	8,05	8,16	7,51	6,81	7,61	6,82	8,07
			7,14	7,69	8,36						
3	15500	16	18,05	18,08	18,72	18,72	19,05	18,03	18,29	16,94	17,52
			18,98	19,13	18,98	17,67	19,06	17,98	19,26		
Celkom	35000	40									

Údaje z tab. 2.1 použijeme na :

- odhad strednej hodnoty nákladovosti a úhrnu nákladov v konečnom základnom súbore,
- vyčíslenie prípustnej chyby odhadu strednej hodnoty nákladovosti a úhrnu nákladov v konečnom základnom súbore s 95% pravdepodobnosťou a
- 95% intervalu spoľahlivosti odhadu strednej hodnoty nákladovosti a úhrnu nákladov v konečnom základnom súbore.

Bodový odhad strednej hodnoty nákladovosti v konečnom základnom súbore je

$$\hat{\mu} = \overline{X}_{str} = \frac{1}{N}\sum_{h=1}^{m} N_h \bar{x}_h = \frac{1}{35000}(409302) = 11{,}69\,\text{EUR},$$

kde $\bar{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$ $\qquad (\bar{x}_1 = 4{,}94,\ \bar{x}_2 = 7{,}56,\ \bar{x}_3 = 18{,}40)$.

Bodový odhad úhrnu nákladov v konečnom základnom súbore je

$$\hat{\tau} = N\overline{X}_{str} = 35000 \cdot 11{,}69 = 409150 \text{ EUR}.$$

Prípustná chyba odhadu strednej hodnoty nákladovosti v konečnom základnom súbore s 95% pravdepodobnosťou je

$$\pm u_{1-\alpha/2}\sqrt{D\left(\overline{X}_{str}\right)} = \pm u_{1-\alpha/2}\frac{1}{N}\sqrt{\sum_{h=1}^{m} N_h^2\left(1-\frac{n_h}{N_h}\right)\frac{s_h^2}{n_h}} = \pm(1{,}96)\frac{1}{35000}\sqrt{9617440} =$$

$$= \pm(1{,}96)\cdot(0{,}0886) = \pm 0{,}174 \text{ EUR}$$

kde $s_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left(x_{hi} - \bar{x}_h\right)^2$ ($s_1^2 = 0{,}0975$, $s_2^2 = 0{,}2114$, $s_3^2 = 0{,}4665$),

Prípustná chyba odhadu úhrnu nákladov v konečnom základnom súbore s 95% pravdepodobnosťou je

$$\pm u_{1-\alpha/2}\sqrt{D\left(N\overline{X}_{str}\right)} = \pm u_{1-\alpha/2}\sqrt{\sum_{h=1}^{m} N_h^2\left(1-\frac{n_h}{N_h}\right)\frac{s_h^2}{n_h}} = \pm(1{,}96)\cdot\sqrt{9617440} = \pm 6078{,}35 \text{ EUR}$$

95%-ný interval spoľahlivosti odhadu strednej hodnoty nákladovosti a úhrnu nákladov v konečnom základnom súbore je

$$\overline{X}_{str} \pm u_{1-\alpha/2}\frac{1}{N}\sqrt{\sum_{h=1}^{m} N_h^2\left(1-\frac{n_h}{N_h}\right)\frac{S_h^2}{n_h}} = 11{,}69 \pm 0{,}174 = \left(11{,}52 \text{ EUR}; 11{,}86 \text{ EUR}\right).$$

95%-ný interval spoľahlivosti odhadu úhrnu nákladov v konečnom základnom súbore je

$$N\overline{X}_{str} \pm u_{1-\alpha/2}\sqrt{\sum_{h=1}^{m} N_h^2\left(1-\frac{n_h}{N_h}\right)\frac{S_h^2}{n_h}} = 409150 \pm 6078{,}35 = \left(403071{,}65 \text{ EUR}; 415228{,}35 \text{ EUR}\right)$$

2.3 Stanovenie rozsahu výberu

Rozsah výberu potrebný k odhadu strednej hodnoty konečného základného súboru (pri výbere bez opakovania), aby prípustná chyba odhadu bola s pravdepodobnosťou (1- α) najviac Δ sa vypočíta

$$n \geq \frac{u_{1-\frac{\alpha}{2}}^2 \sum_{h=1}^{m} N^2\sigma_h^2 / W_h}{N^2\Delta^2 + u_{1-\frac{\alpha}{2}}^2 \sum_{h=1}^{m} N_h\sigma_h^2}, \text{ kde } W_h = \frac{N_h}{N}.$$

V prípade, keď nepoznáme rozptyl σ_h^2 v jednotlivých stratách základného súboru, nahradíme ho jeho odhadom s_h^2 z pilotnej štúdie alebo na základe historických údajov z podobnej štúdie.

Rozsah výberu potrebný k odhadu úhrnu konečného základného súboru , aby prípustná chyba odhadu bola s pravdepodobnosťou (1- α) najviac Δ, sa vypočíta podľa vzťahu:

$$n \geq \frac{u_{1-\frac{\alpha}{2}}^2 \sum_{h=1}^{m} N^2\sigma_h^2 / W_h}{N^2\frac{\Delta^2}{N^2} + u_{1-\frac{\alpha}{2}}^2 \sum_{h=1}^{m} N_h\sigma_h^2}$$

Príklad 2.2 Úlohy v príklade 2.1 rozšírime o stanovenie minimálneho rozsahu výberu, keď podnik požaduje maximálnu prípustnú chybu Δ =0,1 EUR pri odhade strednej hodnoty nákladovosti v základnom súbore s pravdepodobnosťou 0,95.
Potom

$$n \geq \frac{u^2_{1-\frac{\alpha}{2}} \sum_{h=1}^{m} N^2 \sigma_h^2 / W_h}{N^2 \Delta^2 + u^2_{1-\frac{\alpha}{2}} \sum_{h=1}^{m} N_h \sigma_h^2} == \frac{5125163653}{12290450} = 417,$$

kde $W_1 = \frac{N_1}{N} = \frac{8900}{35000} = 0{,}254$, $W_2 = \frac{N_2}{N} = \frac{106000}{35000} = 0{,}303$, $W_3 = \frac{N_3}{N} = \frac{15500}{35000} = 0{,}443$.
Keď by podnik požadoval maximálnu prípustnú chybu 10 centov pri odhade strednej hodnoty nákladovosti v základnom súbore s pravdepodobnosťou 0,95, rozsah výberov v jednotlivých stratách pri stratifikovanom výbere s konštantným výberovým pomerom bude $n_h = \frac{n}{N} N_h$, t.j. $n_1 = 106$, $n_2 = 126$, $n_3 = 185$.

3. Záver

Stratifikované náhodné vyberanie je vhodnejšie ako jednoduché náhodné vyberanie, keď môžeme základný súbor rozdeliť na viaceré nehomogénne skupiny (stratá) tak, že výberové jednotky základného súboru v každom strate sú si podobné, ale za jednotlivé stratá sa líšia. Každé stratum vystupuje samostatne ako základný súbor (v rámci celku ako podsúbor) a jednoduchý náhodný výber je robený z týchto podsúborov (strat).

Tento príspevok vznikol s prispením grantovej agentúry VEGA v rámci projektu číslo 1/0437/08 Kvantatitatívne metódy v stratégii šesť sigma a projektu číslo 1/3182/06 Zlepšovanie kvality produkcie strojárenských výrobkov pomocou štatistických metód.

Literatúra

[1] BHISHAM C. GUPTA - H. FRED WALKER: Statistical Quality Control for the Six Sigma Green Belt. ASQ, Quality Press, Milwaukee 53203, 2007, H 1277.
[2] COCHRAN, W. G.: Sampling Techniques.: J. Wiley and Sons. New York 1977. ISBN 0-471-16240-X.
[3] ČERMÁK - V., VRABEC, M.: Teorie výběrových šetření, Část 3., Vysoká škola ekonomická, Praha, 1999, ISBN 80-245-0003-5.
[4] TEREK M. - HRNČIAROVÁ Ľ.: Výberové skúmanie, Ekonóm, Bratislava, 2008, ISBN 978-80-225-2440-7.

Adresa autora:

Hrnčiarová Ľubica, doc. Ing. PhD.
Katedra štatistiky FHI EU v Bratislave,
Dolnozemská cesta 1
852 35 Bratislava
hrnciaro@euba.sk

Modelovanie intenzity sťahovania SKK z obehu
Modelling Intensity Removing SKK from Currency Circulation

Jozef Chajdiak

Abstract: Paper contains a model for removing balance in hand from currency circulation, estimation of its parameters, the prognosis of future development and study about aspects of coming development.

Key words: Time series, balance in hand in SKK, exponential regression model, power regression model, prognosis of coming development.

Kľúčové slová: Časový rad, peňažná hotovosť v SKK, exponenciálny regresný model, mocninový regresný model, prognóza budúceho vývoja.

1. Úvod

Od 1.1.2009 sa oficiálnym platidlom v Slovenskej republike stala jednotná mena Európskej únie Euro. V čase od 1.1. do 16.1.2009 bol duálny obeh SKK aj EUR. Od 17.1.2009 sa môže v platobnom styku používať len euro. Jedným z aspektov pre Slovenskú republiku principiálnej, strategickej, historickej udalosti je postupná výmena hotovosti z SKK na EUR. Národná banka Slovenska na svojej webovskej stránke zverejňuje stav peňažnej hotovosti k jednotlivým dňom periódy duálneho obehu. Podľa údajov z 20.1.2009 stavy peňazí sú uvedené v tab.1.

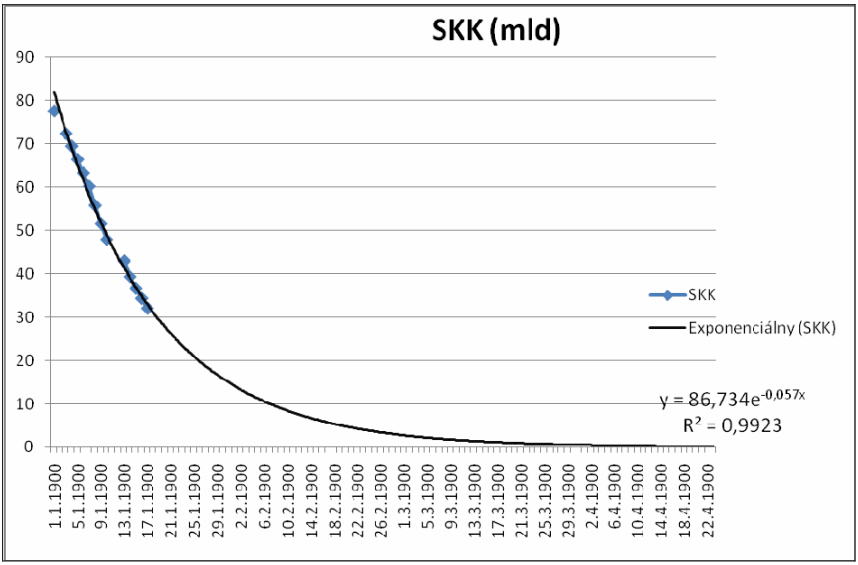
Tabuľka 1: Stav hodnoty slovenských korún v obehu

Dátum	Hodnota slovenských korún v obehu *Sk	Hodnota slovenských korún v obehu * mld Sk
31.12.2008	77642407277	77,64
1.1.2009		
2.1.2009	72452577284	72,45
3.1.2009	69546245662	69,55
4.1.2009	66452416817	66,45
5.1.2009	63358083103	63,36
6.1.2009	60238140921	60,24
7.1.2009	55824694277	55,82
8.1.2009	51638755947	51,64
9.1.2009	47961062190	47,96
10.1.2009		
11.1.2009		
12.1.2009	43033842901	43,03
13.1.2009	39311419089	39,31
14.1.2009	36704013129	36,70
15.1.2009	34366738066	34,37
16.1.2009	32070738824	32,07

Úlohou je odhadnúť budúci vývoj časového radu zostávajúcej hotovosti v obehu.

2. Exponenciálny regresný model

Vývoj časového radu znázorníme čiarovým grafom. V jeho rámci aktivujeme podsystém Pridať trendovú čiaru. Aktivujeme Exponenciálny model, požaduje prognózu na 90 dní dopredu, požadujeme výstup hodnoty R^2 a odhad parametrov exponenciálneho modelu. Výstup je na obr.1.



Obrázok 1: Exponenciálny regresný model stavu hotovosti v mld.SKK

Hodnota R^2 = 0,9923 je veľmi vysoká (hoci čiastočne to súvisí aj s malým počtom pozorovaní). Odhadnutý regresný exponenciálny model má tvar

$$SKKHATt = 86{,}734 * e^{-0{,}057 * t}$$

K zostávajúcemu objemu 1 mld. SKK sa podľa tohto modelu dostaneme 19.3.2009 (x=79); x=1 dňa 31.12.2008).

3. Mocninový regresný model

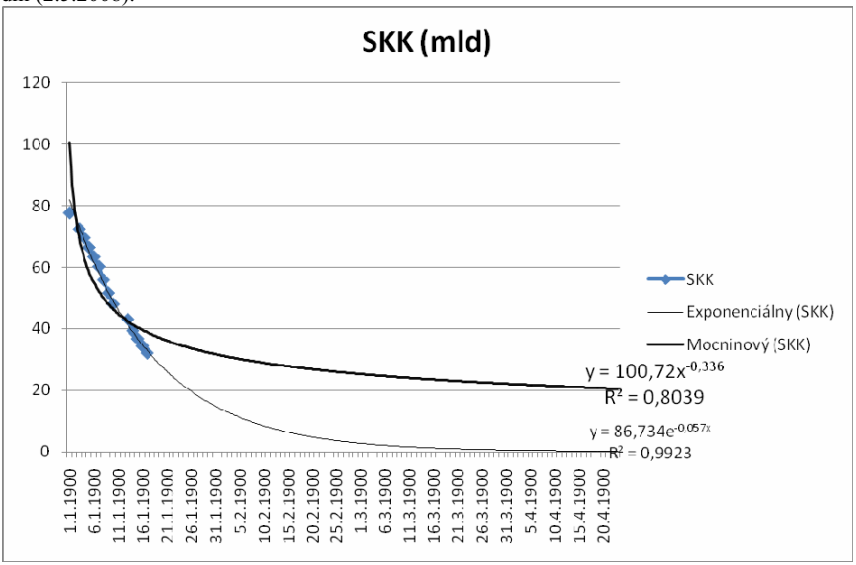
Vývoj časového radu znázorníme čiarovým grafom. V jeho rámci aktivujeme podsystém Pridať trendovú čiaru. Aktivujeme Silový (mocninový) model, požaduje prognózu na 90 dní dopredu, požadujeme výstup hodnoty R^2 a odhad parametrov mocninového modelu. Výstup je na obr.2.

Hodnota R^2 = 0,8039 je tiež dosť vysoká. Odhadnutý regresný mocninový model má tvar

$$SKKHATt = 100{,}72 * t^{-0{,}336}$$

K zostávajúcemu objemu 1 mld. SKK sa podľa tohto modelu dostaneme o 915491 dní (2508 rokov), k zostávajúcemu objemu 10 mld. SKK sa podľa tohto modelu dostaneme o 967 dní

(2,65 roku), k zostávajúcemu objemu 20 mld. SKK sa podľa tohto modelu dostaneme o 123 dní (2.5.2008).



Obrázok 2: Mocninový regresný model stavu hotovosti v mld. SKK

4. Zamyslenie sa nad kvalitou modelu

Expertné očakávania nevymeneného objemu SKK sa pohybujú medzi 10-20 mld. SKK Podľa exponenciálneho modelu na 1 mld. SKK sme už v marci 2009, čiže medzi expertnými a modelovými odhadmi je zrejmý nesúlad. Mocninový model je bližšie k expertným očakávaniam – po dvoch rokoch očakáva stav 11 mld. SKK (čo predstavuje mimoriadny bezplatný úver pre sektor goverment vo výške 11 mld. SKK na 2 roky).

Jedným z aspektov tohto výsledku je fakt, že údaje vyjadrujú jednu reálnu situáciu (duálny obeh od 1.1. po 16.1.2009) a odhadnutý model sme použili na odhad vývoja stavu peňažnej hotovosti v SKK pri situácii, keď sa už hotovosťou v SKK nedá platiť a dá sa len vymeniť v bankách.

5. Literatúra

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

[2] CHAJDIAK, J. 2003. Štatistické úlohy a ich riešenie v Exceli. Bratislava: Statis 2005. 262 s.

[3] CHAJDIAK, J.: 2009. Štatistika v Exceli 2007. Bratislava: Statis 2009. 312 s. ISBN 978-80-85659-49-8.

Adresa autora:

Jozef Chajdiak, Doc., Ing., CSc.
Ústav manažmentu STU
810 00 Bratislava
chajdiak@statis.biz

Je európsky trh s kokaínom jednotný?
Is the European Cocaine Market Homogenous?

Jaroslav Ivor, Beáta Stehlíková, Anna Tirpáková, Dagmar Markechová

Abstract: We analyze cocaine retail and whole sale prices in Europe with the help of statistic methods. We look for spatial market integration.

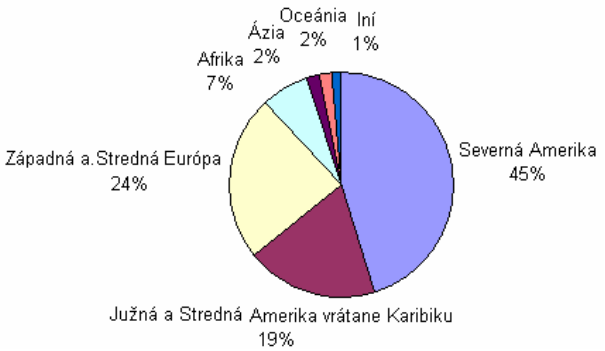
Key words: drug market, market integration, time series

Kľúčové slová: trh s drogami, integrácia trhu, časové rady

1. Úvod

Antonio Maria Costa, výkonný riaditeľ Úradu pre drogy a kriminalitu OSN (UNODC) predstavil 26. júna 2008 v New Yorku pri príležitosti Medzinárodného dňa boja proti zneužívaniu drog a ilegálnemu obchodovaniu novú svetovú správu o drogách „2008 World Drug Report**“** [8] . Správa konštatuje že každý dvadsiaty človek na svete vo veku od 15 do 64 rokov užil drogu najmenej raz za posledných 12 mesiacov**.**

Kokaín je po kanabise druhou najčastejšie užívanou nezákonnou drogou aj v Európe. V užívaní kokaínu existuje významný rozdiel medzi európskymi krajinami. Užívanie kokaínu sa sústreďuje v niekoľkých krajinách. V krajinách severnej a strednej Európy prevláda užívanie amfetamínov a v krajinách na západe a juhu Európy prevažuje kokaín. Krajiny s najvyššou prevalenciou sú Spojené kráľovstvo (7,7 %), Španielsko (7,0 %), Taliansko (6,6 %), Írsko (5,3 %). Najnižšiu prevalenciu užívania kokaínu vykazujú štáty Rumunsko, Malta, Litva (0,4 %) a Grécko (0,7 %). Prieskumy zaznamenali výrazný nárast užívania medzi mladými ľuďmi od polovice deväťdesiatych rokov. Najnovšie údaje naďalej poukazujú na celkový nárast v užívaní kokaínu v Európe.



Zdroj: [8]

Obrázok 1: Ročná prevalencia užívania kokaínu v jednotlivých častiach sveta (2006/2007)

Kokaín je najčastejšie obchodovanou drogou na svete po trávovom kanabise a kanabisovej živici. Pestovanie kríkov koky, zdroja kokaínu, sa sústreďuje do Kolumbie, za ňou nasleduje Peru a Bolívia. Kokaín sa do Európy pašuje cez Brazíliu, Ekvádor, Venezuelu a Karibik. V posledných rokoch sa výrazne zintenzívňuje transport cez krajiny západnej Afriky.

Hlavnými miestami vstupu kokaínu do Európy sú Španielsko a od roku 2005 tiež Portugalsko. Europol [10] uvádza Holandsko a Francúzsko ako hlavné tranzitné krajiny pre ďalšiu distribúciu kokaínu v Európe. Kokaín prichádza do Európy letecky tiež z Belgicka, Talianska, Francúzska, Spojeného kráľovstva a Nemecka [9]. Na vývoj nových obchodných ciest poukazujú správy o dovoze kokaínu cez Bulharsko, Estónsko, Lotyšsko, Litva, Rumunsko, Rusko, konštatuje sa v [9].

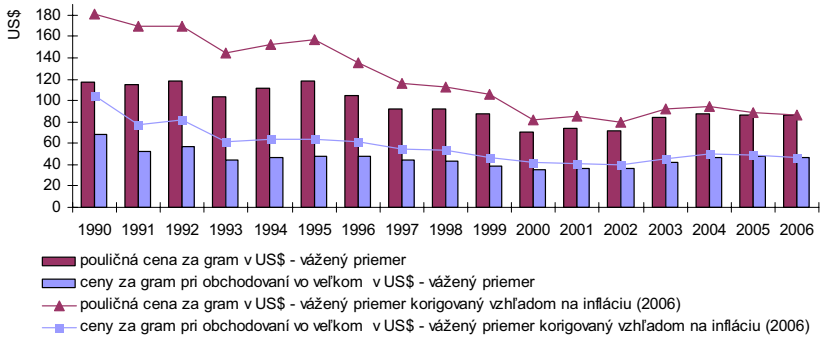
Zachytenia drog sa považuje za nepriamy ukazovateľ ponuky, obchodných trás a dostupnosti. Odráža tiež prostriedky a stratégie presadzovania práva ako aj, zraniteľnosť priekupníkov. V roku 2006 sa na celosvetovej úrovni znížil počet zachytení kokaínu na 706 ton. Počet zachytení kokaínu v Európe sa však zvýšil na 72 700 prípadov a odhalené množstvo na 121 ton [8]. Na Španielsko pripadá 58% všetkých zachytení a 41% množstva, na Portugalsko 28 percent zachyteného množstva kokaínu.

Tabuľka 1: Zachytenia kokaínu v Európe v roku 2006

Štát	Zachytené množstvo (t)	Podiel z celkového zachyteného kokaínu v Európe (%)
Španielsko	49,7	40,9
Portugalsko	34,5	28,4
Holandsko	10,6	8,7
Francúzsko	10,2	8,4
Taliansko	4,6	3,8
Belgicko	3,9	3,2
Anglicko	3,8	3,1
Nemecko	1,7	1,4
Švédsko	1,4	1,1
Švajčiarsko	0,4	0,3
Poľsko	0,3	0,2
Bulharsko	0,1	0,1
Turecko	0,1	0,1
Dánsko	0,1	0,1
Rakúsko	0,1	0,1
iné	0,2	0,2

Zdroj: [8]

Údaje o pouličných cenách sa ťažko zhromažďujú a interpretujú. Čistota, množstvo a druh kupovanej látky ovplyvňujú cenu práve tak, ako aj geografické faktory, napríklad bývanie vo veľkom meste alebo na pravidelnej tranzitnej drogovej trase. Ceny drog sa značne líšia aj medzi jednotlivými krajinami a podliehajú časovým zmenám, ktoré odrážajú prerušenia ponuky, uvádza sa v [9] Prehľad priemerných veľkoobchodných a pouličných cien kokaínu v Európe za roky 1990 až 2006 je znázornený na obrázku 2.



Zdroj: [8] a vlastné zobrazenie

Obrázok 2: Priemerné ceny kokaínu v Európe

2. Materiál a metódy

Cenová korelácia sa stala štandardným nástrojom európskeho súťažného práva [1]. Analýza cenovej korelácie je metóda, ktorá slúži k stanoveniu relevantného trhu. Pomocou štatistických metód sa zisťuje podobnosť vo vývoji cien skúmaných výrobkov. Ekonomickým základom je predpoklad, že ceny výrobkov na zhodnom trhu sa vyvíjajú podobným spôsobom. Vzťah medzi cenami v priestore predstavuje dôležitý indikátor integrácie trhu. Stigler [7] definoval trh ako územie, v rámci ktorého cena tovaru má tendenciu byť rovnaká, ak neberieme do úvahy náklady na prepravu. To znamená, že keď analyzujeme vzťahy medzi cenami, korelácia medzi cenami je často používaná k determinácii, či dve geografické územia sú tým istým ekonomickým trhom.

Vývoj cien predstavuje zo štatistického hľadiska časový rad. Pri skúmaní závislostí medzi časovými radmi, najčastejšie vychádzame z predpokladu, že sa dajú popísať aditívnym modelom. V prípade, že chceme skúmať, či medzi časovými radmi existuje určitý príčinný vzťah, nie je postačujúce skúmať iba celkovú vývojovú tendenciu resp. sezónne kolísanie. A to z toho dôvodu, že tieto faktory môžu mať veľmi podobný priebeh v dôsledku multikolinearity. Je preto potrebné skúmať, či existuje vzťah medzi náhodnými zložkami skúmaných resp. pozorovaných časových radoch. Ak existuje závislosť medzi náhodnými zložkami, je na mieste predpokladať, že existuje skutočná príčinná závislosť medzi sledovanými časovými radmi [6]. Dôležitá je voľba trendových funkcií. V prípade zle vyrovnávajúcej trendovej funkcie môže dôjsť ku skresleniu skutočnej závislosti medzi časovými radmi. Iný spôsob merania tesnosti závislosti medzi časovými radmi je metóda diferencií. 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 $\Delta x_t = x_t - x_{t-1}$ [3].

Stacionaritu časového radu možno testovať pomocou Augmented Dickeyovho-Fullerovho testu [4].

Koeficienty korelácie medzi diferenciami analyzovaných časových radov predstavujú miery tesnosti medzi reziduálnymi zložkami týchto radov. Koeficient korelácie medzi diferenciami dvoch časových radov vypočítame podľa vzťahu:

$$r_{\Delta x_t \Delta y_t} = \frac{\frac{\sum \Delta x_t \Delta y_t}{n} - \frac{\sum \Delta x_t}{n} \cdot \frac{\sum \Delta y_t}{n}}{\sqrt{\left\{\frac{\sum \Delta^2 x_t}{n} - \left(\frac{\sum x_t}{n}\right)^2\right\}\left\{\frac{\sum \Delta^2 y_t}{n} - \left(\frac{\sum \Delta y_t}{n}\right)^2\right\}}}.$$

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.

Boissseleau – Hewicker [2] uvádzajú, že pre hodnotu koeficienta korelácie, ktorá je väčšia ako 0,5, sa jedná o medzinárodnú integráciu trhu.

Nech $r(x,y)$ je korelačný koeficient medzi hodnotami časových radov x a y. Hodnota

$$d(x, y) = \sqrt{2(1 - r(x, y))}$$

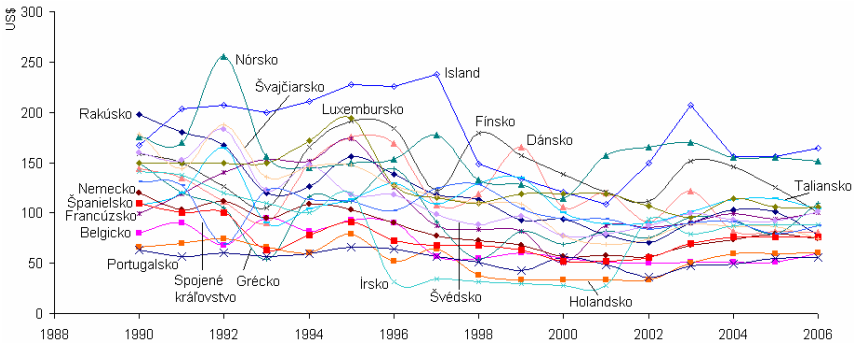
je metrika, ktorá vhodne popisuje podobnosť časových radov [5].

Popísané štatistické metódy boli použité pre overenie integrácie trhu s kokaínom. V príspevku analyzujeme údaje o pouličných cenách kokaínu a tiež ceny pri obchodovaní vo veľkom v európskych štátoch OECD v rokoch 1990-2006. Údaje boli čerpané z World drug reports [8].

3. Výsledky a diskusia

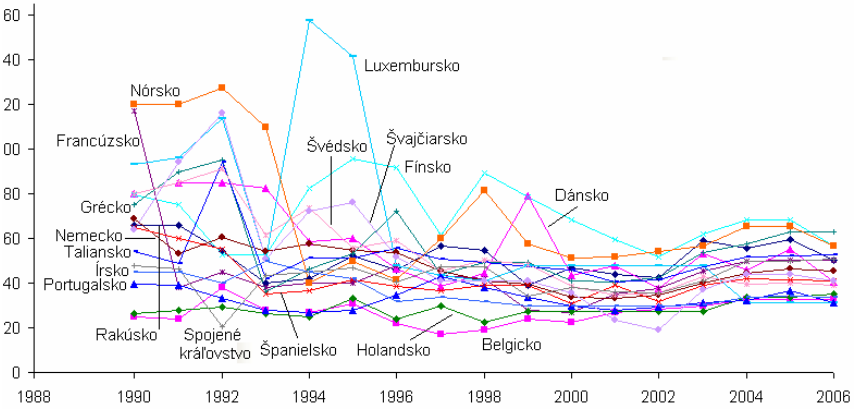
V roku 2006 priemerná pouličná cena kokaínu v celej EÚ veľmi kolísala, od 56 EUR za gram v Portugalsku a 60 EUR v Belgicku a Holandsku, po viac ako 100 EUR za gram vo Švédsku, Taliansku, Luxembursku, Grécku, viac ako 150 EUR v Nórsku a na Islande. Priemerné ceny kokaínu korigované vzhľadom na infláciu ukázali celkový klesajúci trend počas obdobia 1999 – 2006 vo všetkých krajinách, ktoré poskytli správy, s výnimkou Luxemburska, kde klesali až do roku 2002 a potom sa zvýšili, a Nórska, kde ceny v roku 2001 strmo stúpli a potom sa stabilizovali.

Priebeh pouličných cien jedného gramu kokaínu a ceny jedného gramu pri obchodovaní vo veľkom v európskych štátoch OECD v rokoch 1990 až 2006 je znázornený na obrázkoch 3 a 4.



Zdroj:[8] a vlastne zobrazenie

Obrázok 3: Pouličné ceny kokaínu v rokoch 1990-2006



Zdroj:[8] a vlastne zobrazenie

Obrázok 4: Ceny kokaínu pri obchodovaní vo veľkom v rokoch 1990-2006

V ďalšej analýze sme pracovali s logaritmami hodnôt. Stacionaritu časových radov údajov o pouličných cenách kokaínu a tiež aj cien pri obchodovaní vo veľkom v európskych štátoch OECD v rokoch 1990-2006 sme overovali pomocou Augmented Dickeyovho-Fullerovho testu. Časové rady oboch typov cien všetkých analyzovaných štátov vykazovali nestacionaritu. Z tohto dôvodu sme vypočítali prvé absolútne diferencie, ktoré použitím uvedeného testu už boli stacionárne (Tabuľka 2a, 2b), pričom kritické hodnota pre α = 0,01 je -2,728252 a pre α = 0,05 je -1,9605026.

Tabuľka 2a: Výsledky Augmented Dickeyovho-Fullerovho testu stacionarity pre prvé absolútne diferencie logaritmov pouličných cien kokaínu

	Štát	Exogenous	Lag Length	t-Statistic	Prob.*
1	Rakúsko	None	0	-2,949136	0,0061
2	Belgicko	None	0	-5,367491	0,0000
3	Dánsko	Constant, Linear Trend	3	-3,862835	0,0509
4	Fínsko	Constant, Linear Trend	2	-5,200937	0,0063
5	Francúzsko	None	0	-7,028790	0,0000
6	Nemecko	None	0	-3,792103	0,0009
7	Grécko	None	0	-5,414679	0,0000
8	Island	None	0	-3,670839	0,0012
9	Taliansko	None	0	-5,241138	0,0000
10	Luxembursko	None	0	-3,889369	0,0008
11	Holandsko	None	0	-5,383952	0,0000
12	Nórsko	None	0	-4,712439	0,0001
13	Portugalsko	None	0	-4,712439	0,0001
14	Španielsko	None	1	-3,168176	0,0040
15	Švédsko	None	0	-6,440156	0,0000
16	Švajčiarsko	None	0	-4,972748	0,0001
17	Anglicko	None	0	-6,168756	0,0000
18	Írsko	None	0	-4,387311	0,0003

Zdroj: vlastné výpočty

Tabuľka 2b: Výsledky Augmented Dickeyovho-Fullerovho testu stacionarity pre prvé absolútne diferencie logaritmov cien kokaínu pri obchodovaní vo veľkom

	Štát	Exogenous	Lag Length	t-Statistic	Prob.*
1	Rakúsko	None	3	-4,435630	0,0004
2	Belgicko	None	0	-4,625730	0,0002
3	Dánsko	None	0	-5.885647	0.0000
4	Fínsko	None	0	-3.927859	0.0007
5	Francúzsko	None	0	-8.592618	0.0000
6	Nemecko	None	0	-4.605024	0.0002
7	Grécko	None	2	-5.668174	0.0000
8	Taliansko	None	0	-7.761611	0.0000
9	Luxembursko	None	1	-4.809597	0.0001
10	Holandsko	None	0	-9.088080	0.0000
11	Nórsko	None	3	-7,178274	0,0000
12	Portugalsko	None	0	-12,40669	0,0001
13	Španielsko	None	0	-4,784173	0,0001
14	Švédsko	None	0	-5,849985	0,0000
15	Švajčiarsko	None	0	-4,258960	0,0003
16	Anglicko	None	0	-5,790366	0,0000
17	Írsko	None	0	-4,649454	0,0002

Zdroj: vlastné výpočty

Na základe výsledkov testovania (tabuľka 2a, 2b) môžeme povedať, že prvé absolútne diferencie logaritmov oboch cien kokaínu pre všetky hodnotené štáty sú stacionárne. Z tohto dôvodu na určenie závislosti časových radov cien kokaínu medzi jednotlivými štátmi je

korektné použiť koeficient korelácie. Koeficienty korelácie medzi časovými radmi prvých absolútnych diferencií sú uvedené v nasledovných maticiach (tabuľka 3a, 3b).

Tabuľka 3a: Koeficienty korelácie pre prvé absolútne diferencie logaritmov pouličných cien kokaínu štátov OECD za roky1990-2006

	Rakúsko	Belgicko	Dánsko	Fínsko	Francúzsko	Nemecko	Grécko	Island	Taliansko	Luxembursko	Holandsko	Nórsko	Portugalsko	Španielsko	Švédsko	Švajčiarsko	Anglicko	Írsko
Rakúsko	1,00	-0,14	0,21	**0,61**	-0,03	**0,54**	0,21	0,12	**0,51**	0,23	0,47	0,12	**0,61**	**0,70**	0,13	0,47	-0,26	0,06
Belgicko	-0,14	1,00	0,23	0,12	0,32	-0,15	-0,02	0,03	-0,33	0,15	-0,10	-0,62	0,02	-0,29	-0,27	-0,38	0,44	0,02
Dánsko	0,21	0,23	1,00	**0,67**	0,38	0,48	**0,66**	0,09	0,38	0,15	-0,12	0,01	0,05	0,45	0,48	0,37	-0,12	-0,19
Fínsko	**0,61**	0,12	**0,67**	1,00	0,13	0,49	0,34	-0,09	0,46	0,17	-0,15	-0,26	0,25	**0,54**	0,25	0,34	-0,05	-0,08
Francúzsko	-0,03	0,32	0,38	0,13	1,00	0,42	0,13	0,04	0,16	0,41	0,20	0,26	-0,24	0,28	0,25	0,28	-0,03	0,24
Nemecko	**0,54**	-0,15	0,48	0,49	0,42	1,00	0,50	0,16	**0,53**	0,27	0,34	0,33	0,29	**0,75**	**0,68**	**0,80**	-0,37	0,14
Grécko	0,21	-0,02	**0,66**	0,34	0,13	0,50	1,00	0,24	0,42	0,10	0,06	0,31	0,12	**0,51**	**0,73**	0,35	-0,36	-0,20
Island	0,12	0,03	0,09	-0,09	0,04	0,16	0,24	1,00	0,02	-0,22	**0,57**	0,31	0,07	0,23	0,29	0,28	-0,04	0,26
Taliansko	**0,51**	-0,33	0,38	0,46	0,16	**0,53**	0,42	0,02	1,00	-0,02	-0,02	0,49	0,02	**0,67**	**0,70**	**0,74**	-0,74	-0,11
Luxembursko	0,23	0,15	0,15	0,17	0,41	0,27	0,10	-0,22	-0,02	1,00	0,36	-0,11	0,10	0,34	-0,04	0,10	-0,01	0,47
Holandsko	0,47	-0,10	-0,12	-0,15	0,20	0,34	0,06	**0,57**	-0,02	0,36	1,00	0,37	0,41	0,45	0,03	0,31	0,02	0,37
Nórsko	0,12	-0,62	0,01	-0,26	0,26	0,33	0,31	0,31	0,49	-0,11	0,37	1,00	-0,04	0,43	**0,57**	**0,56**	-0,67	0,04
Portugalsko	**0,61**	0,02	0,05	0,25	-0,24	0,29	0,12	0,07	0,02	0,10	0,41	-0,04	1,00	0,19	-0,08	0,06	-0,08	-0,32
Španielsko	**0,70**	-0,29	0,45	**0,54**	0,28	**0,75**	**0,51**	0,23	**0,67**	0,34	0,45	0,43	0,19	1,00	**0,61**	**0,75**	-0,39	0,35
Švédsko	0,13	-0,27	0,48	0,25	0,25	**0,68**	**0,73**	0,29	**0,70**	-0,04	0,03	**0,57**	-0,08	**0,61**	1,00	**0,71**	-0,66	0,03
Švajčiarsko	0,47	-0,38	0,37	0,34	0,28	**0,80**	0,35	0,28	**0,74**	0,10	0,31	**0,56**	0,06	**0,75**	**0,71**	1,00	-0,59	0,21
Anglicko	-0,26	0,44	-0,12	-0,05	-0,03	-0,37	-0,36	-0,04	-0,74	-0,01	0,02	-0,67	-0,08	-0,39	-0,66	-0,59	1,00	0,01
Írsko	0,06	0,02	-0,19	-0,08	0,24	0,14	-0,20	0,26	-0,11	0,47	0,37	0,04	-0,32	0,35	0,03	0,21	0,01	1,00

Zdroj: Vlastné výpočty

V zmysle Boissseleauovho – Hewickerovho vymedzenia ([2]) v prípade Rakúska dostávame, že rakúsky trh kokaínu je integrovaný s trhmi Fínska, Nemecka, Talianska, Portugalska a Španielska. Dánsky trh je integrovaný s trhmi Fínska a Grécka a fínsky trh s trhmi Rakúska, Dánska a Španielska. Nemecký trh kokaínu je integrovaný s rakúskym, talianskym, španielskym, švédskym a švajčiarskym trhom. V prípade Grécka vidíme, že grécky trh je integrovaný s trhmi Dánska, Španielska a Švédska. Pouličné ceny kokaínu sa v Taliansku vyvíjali v hodnotenom období podobne ako ceny v Rakúsku, Nemecku, Španielsku, Švédsku a Švajčiarsku.

Holandský trh kokaínu je integrovaný s trhom Islandu. Trh Nórska je integrovaný s belgickým, švédskym a švajčiarskym trhom; portugalský trh s trhom Rakúska. Španielsky trh je integrovaný s trhmi Rakúska, Fínska, Nemecka, Grécka, Talianska, Švédska a s trhom Švajčiarska. Trh Švédska je integrovaný s nemeckým, gréckym, talianskym, nórskym, španielskym a švajčiarskym trhom. A nakoniec trh Švajčiarska je integrovaný s trhmi Nemecka, Talianska, Nórska, Španielska a Švédska.

Tabuľka 3b: Koeficienty korelácie pre prvé absolútne diferencie logaritmov cien pri obchodovaní vo veľkom štátov OECD za roky1990-2006

	Rakúsko	Belgicko	Dánsko	Fínsko	Francúzsko	Nemecko	Grécko	Taliansko	Luxembursko	Holandsko	Nórsko	Portugalsko	Španielsko	Švédsko	Švajčiarsko	Anglicko	Írsko
Rakúsko	1,00	-0,13	-0,17	0,22	0,14	-0,02	0,24	0,18	0,24	0,20	0,26	**0,52**	0,37	0,04	0,45	0,19	0,01
Belgicko	-0,13	1,00	0,41	-0,04	0,12	0,45	0,41	**0,63**	0,46	0,34	0,11	-0,48	0,45	0,33	0,47	-0,71	-0,17
Dánsko	-0,17	0,41	1,00	0,10	-0,23	0,17	0,11	-0,04	0,02	0,18	0,18	-0,08	0,38	0,19	0,24	-0,21	0,07
Fínsko	0,22	-0,04	0,10	1,00	0,08	0,21	0,37	-0,06	0,26	-0,34	-0,35	-0,04	0,31	0,36	0,35	0,31	-0,11
Francúzsko	0,14	0,12	-0,23	0,08	1,00	**0,69**	0,08	0,29	-0,08	-0,20	0,03	0,11	0,26	0,05	-0,17	-0,06	-0,18
Nemecko	-0,02	0,45	0,17	0,21	**0,69**	1,00	0,40	**0,55**	0,17	-0,02	-0,19	0,03	0,38	0,48	0,34	-0,31	-0,19
Grécko	0,24	0,41	0,11	0,37	0,08	0,40	1,00	**0,72**	0,33	-0,11	-0,17	0,22	**0,66**	**0,79**	**0,69**	-0,57	-0,73
Taliansko	0,18	**0,63**	-0,04	-0,06	0,29	**0,55**	**0,72**	1,00	0,43	0,11	-0,05	0,08	0,46	**0,63**	**0,65**	-0,83	-0,64
Luxembursko	0,24	0,46	0,02	0,26	-0,08	0,17	0,33	0,43	1,00	0,25	-0,48	-0,27	0,32	0,42	**0,58**	-0,31	-0,10
Holandsko	0,20	0,34	0,18	-0,34	-0,20	-0,02	-0,11	0,11	0,25	1,00	0,22	0,05	0,21	-0,39	0,25	-0,08	0,35
Nórsko	0,26	0,11	0,18	-0,35	0,03	-0,19	-0,17	-0,05	-0,48	0,22	1,00	0,23	0,09	-0,31	-0,10	0,02	0,25
Portugalsko	**0,52**	-0,48	-0,08	-0,04	0,11	0,03	0,22	0,08	-0,27	0,05	0,23	1,00	0,22	0,06	0,07	0,13	-0,20
Španielsko	0,37	0,45	0,38	0,31	0,26	0,38	**0,66**	0,46	0,32	0,21	0,09	0,22	1,00	0,43	**0,55**	-0,31	-0,32
Švédsko	0,04	0,33	0,19	0,36	0,05	0,48	**0,79**	**0,63**	0,42	-0,39	-0,31	0,06	0,43	1,00	**0,67**	-0,51	-0,55
Švajčiarsko	0,45	0,47	0,24	0,35	-0,17	0,34	**0,69**	**0,65**	**0,58**	0,25	-0,10	0,07	**0,55**	**0,67**	1,00	-0,38	-0,23
Anglicko	0,19	-0,71	-0,21	0,31	-0,06	-0,31	-0,57	-0,83	-0,31	-0,08	0,02	0,13	-0,31	-0,51	-0,38	1,00	**0,68**
Írsko	0,01	-0,17	0,07	-0,11	-0,18	-0,19	-0,73	-0,64	-0,10	0,35	0,25	-0,20	-0,32	-0,55	-0,23	**0,68**	1,00

Zdroj: Vlastné výpočty

V zmysle Boissseleau – Hewicker vymedzenia ([2]) sa v prípade Belgicka jedná o medzinárodnú integráciu belgického trhu kokaínu spolu s trhom Talianska. Trh Francúzska je integrovaný s trhom Nemecka. Nemecký trh je integrovaný s francúzskym a talianskym trhom. Ceny kokaínu na trhu Grécka pri predaji vo veľkom sa v analyzovanom období vyvíjali podobne ako v Taliansku, Španielsku, Švédsku a Švajčiarsku. V prípade talianskeho trhu je možné uvažovať v uvedenom zmysle o jeho medzinárodnej integrácii s trhmi Belgicka, Nemecka, Grécka, Švédska a Švajčiarska. Luxemburský trh s kokaínom je integrovaný so švajčiarskym trhom. Portugalský trh je integrovaný s trhom Rakúska a španielsky trh je integrovaný s trhmi Grécka a Švajčiarska.

Ceny kokaínu na trhu vo Švédsku pri predaji vo veľkom sa vyvíjali podobne ako na trhoch v Grécku, Taliansku a Švajčiarsku a ceny kokaínu na trhu vo Švajčiarsku sa vyvíjali podobne ako na trhoch v Grécku, Taliansku, Luxembursku, Španielsku a Švédsku. Trh s kokaínom v Anglicku je integrovaný s írskym trhom.

Výsledky cenovej korelácie korešpondujú s výsledkami analýzy podobnosti časových radov.

4. Záver

V príspevku sme sledovali ceny kokaínu na európskom trhu pri pouličnom predaji a predaji vo veľkom.. Sledovali sme vzťah medzi cenami v priestore ako dôležitého indikátora integrácie trhu. Pre analyzovanie vzťahov medzi cenami sme použili koreláciu, ktorá v prípade časových radov očistených o trendovú a cyklickú zložku nám môže dať odpoveď na otázku, či dve geografické územia sú tým istým ekonomickým trhom. V našom prípade sa potvrdilo, že európsky trh s kokaínom štátov OECD nie je integrovaný ako celok. Výsledky však poukazujú na integráciu trhov niektorých štátov OECD.

5. Literatúra

[1]BISHOP, S. – WALKER, M.: The Economics of EC Competition Law. London: Sweet & Maxwell Ltd, 2002, 443 s.

[2] BOISSEELEAU, F. – HEWICKER, C.: European Electricity Market Design and its Impact on market Integration.

[3] CIPRA, T.: Statistická analýza časových řad s aplikacemi v ekonomii, SNTL, Praha, 1986

[4]MACHOVÁ, K.: Menová politika a hospodársky cyklus vo vybraných nových členských krajinách EÚ. Diplomová práca, KU Bratislava, 2008

[5]MAHARAJ, E. A.: Clusters of time series. Journal of Classification, 17, 2000, 297- 314

[6]SEGER, J.: Statistické metódy pro ekonomy průmyslu. SNTL, Praha 1988, 512 s.

[7]STIGLER, G. J.: Does economics have a useful past? Hist. Polit. Econ. 1,217-230

[8]World Drug Report, ISBN 0-19-829320-8, http://www.unodc.org/unodc/en/data-and-analysis/WDR-2007.html

[9]Stav drogovej problematiky v Európe. Úrad pre vydávanie úradných publikácií Európskych spoločenstiev, Luxemburg 2006. ISBN 92-9168-261-6 Európske monitorovacie centrum pre drogy a drogovú závislosť, 2006.

[10] Europol. Project COLA: European Union cocaine situation report 2007, Europol, Haag.

Adresy autorov:

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

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

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

Dagmar Markechová, Doc. RNDr., CSc.
Katedra Matematiky UKF
Tr. A. Hlinku 1
949 01 Nitra1
dmarkechova@ukf.sk

Matematické modelování věkových struktur historických populací
Mathematical models of age structure of historical populations

Eva Kačerová, Jiří Henzler

Abstract: The List of Inhabitants according to Faith of 1651 captured the situation immediately following the end of the Thirty Years War. The first column in the List contained the name of the person concerned, the second one is age and the profession and religion are filed in the last columns. On the Chrudim area children are recorded systematically only from the age of 11. This article is focused on age distribution and mathematical modelling of age distribution. This article came into being within the framework of the research project IGA VŠE 15/08.

Key words: Historical demography, Chrudim area, age structure, mathematical modelling

Klíčová slova: Historická demografie, Chrudimský kraj, věková struktura, matematické modelování

1. Úvod

Základním pramenem pro studium složení obyvatelstva podle věku a pohlaví v Chrudimském kraji v polovině 17. století je Soupis poddaných podle víry z roku 1651 (Pazderová, 2002), jehož vznik souvisí s rekatolizačními snahami po třicetileté válce. Impulsem ke zhotovení Soupisu byl patent českých místodržících ze dne 16. listopadu 1650, který byl vydán 4. února 1651. Konstatovala se v něm nedokonalost dosavadních soupisů nekatolických poddaných a nařizovalo se všem vrchnostem, aby podle přiloženého formuláře uvedly „všechny lidi poddané obojího pohlaví v jistou a hodnověrnou specifikaci“, již měli krajští hejtmané odeslat do šesti neděl české kanceláři.

2. Obyvatelstvo Chrudimského kraje

Podle Soupisu žilo na Chrudimsku 46 626 osob, z toho 24 860 žen. U 2 104 osob nebyl údaj o věku zapsán. Děti mladší 12 let byly evidovány zřídka. Na žádném z panství Chrudimského kraje nebyla evidence dětí „předzpovědního“ věku úplná, a tak k odhadu počtu dětí mladších 10 resp. 12 let by musela být užita analogie s jiným krajem (není předmětem tohoto článku).

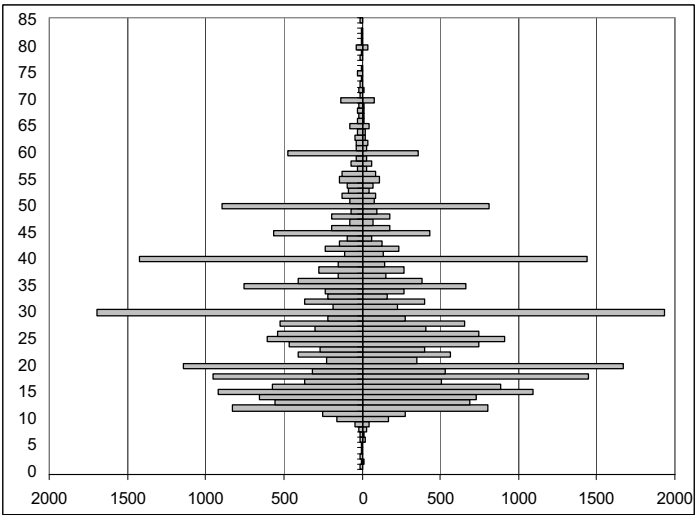
Pro studium demografické struktury je velice důležitý údaj o věku, který je čtvrtou kategorií Soupisu. Přesný věk nebyl příliš důležitý pro účel, za kterým Soupis vznikal, a hrál spíše podružnou roli. I přes zjevné nepřesnosti a zaokrouhlování představuje cennou informaci (obr.1). Písaři měli tendenci věk těch, kteří jej neznali přesně, zaokrouhlovat na násobky 10 či 5. Větší oblibu než lichá čísla měla čísla sudá. Ve věku mladším 30 let by se dalo hovořit i o zaokrouhlování vlivem šedesátkové soustavy. Zajímavá je také souvislost věku s některými sociálními kategoriemi. Vdovám podruhyním se často připisoval věk 40 let, pokud byly starší tak 60 let. U „žen samotných“ majících děti, tedy pravděpodobně svobodných matek, zase často najdeme věk 30 let a podobně. Míru zkreslení lze měřit indexem věkové kumulace *ik*:

$$ik^p = (5 * \sum_{0}^{7} S^p_{25+5x}) / (\sum_{23}^{62} S^p_x), \qquad (1)$$

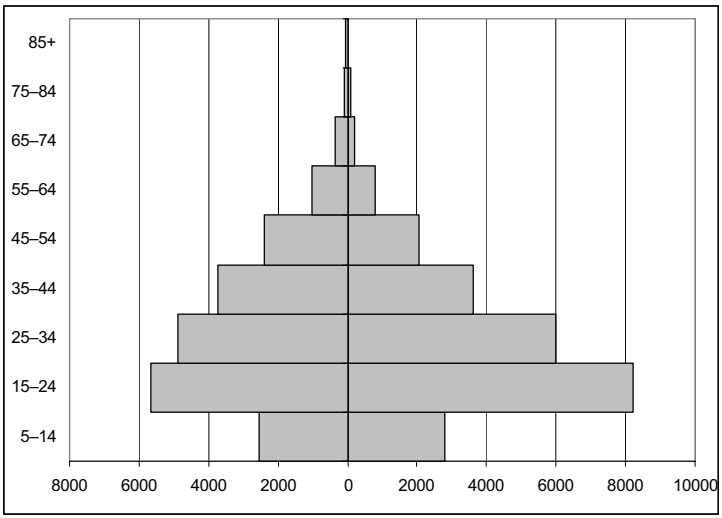
kde S_x je představuje počet osob v dané věkové skupině a horní index p je pohlaví.

Zaokrouhlování věku se negativně promítne do výsledků studia věkové struktury obyvatelstva. Toto zkreslení lze zmenšit používáním desetiletých věkových intervalů, v nichž

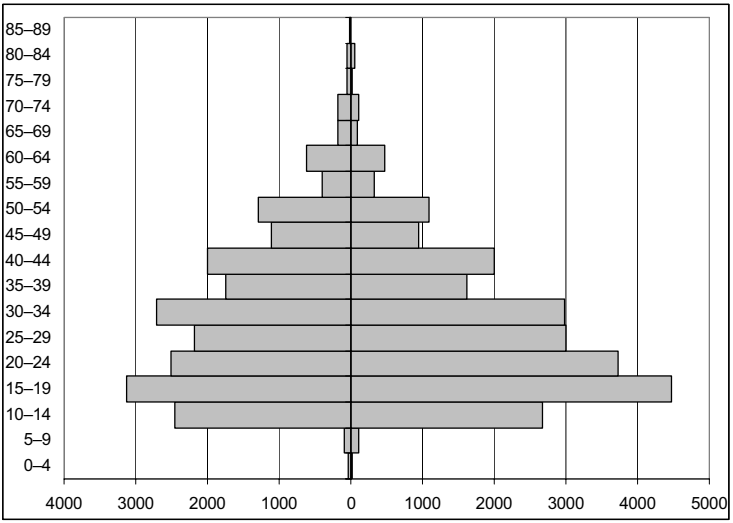
bude nejfrekventovanější hodnota vždy uprostřed, tedy 5–14, 15–24, atd. (obr. 2). Pro možnost srovnání s dnešní statistikou i ostatními autory zabývajícími se rozborem Soupisu poddaných podle víry z roku 1651 je potřeba zachovat také obvyklé věkové intervaly: 0–4, 5–9, 10–14, atd. (obr. 3). V některých studiích se ještě můžeme setkat s desetiletými intervaly 0–9, 10–19, atd.. Volba věkových intervalů může ovlivnit význam zaokrouhlování ve výsledné věkové struktuře. V každém případě však vede ke ztrátě informace.



Obrázek 1: Věková struktura obyvatel Chrudimského kraje dle Soupisu poddaných podle víry z roku 1651



Obrázek 2: Věková struktura obyvatel Chrudimského kraje dle Soupisu poddaných podle víry z roku 1651 (desetileté intervaly)



Obrázek 3: Věková struktura obyvatel Chrudimského kraje dle Soupisu poddaných podle víry z roku 1651 (pětileté intervaly)

3. Matematické modely

Druhou možností je vytvořit jednoduchý matematický model předpokládající, že zaokrouhlování vykazuje určitou zákonitost. Pokud teoretická data získaná na základě takového modelu nebudou vykazovat systematickou chybu, můžeme na základě použitého modelu vyslovit hypotézu o možném způsobu zaokrouhlování.

V této stati autoři předkládají dva matematické modely. V obou se předpokládá, že na celé desítky se zaokrouhluje věk nejvýše o 4 roky větší a nejvýše o 4 roky menší, na hodnoty zakončené 5 se zaokrouhluje věk nejvýše o 2 roky větší a nejvýše o 2 roky menší, na sudá čísla se zaokrouhluje věk nejvýše o 1 rok větší a nejvýše o 1 rok menší. Zaokrouhlování na čísla dělitelná 6 nebylo v těchto modelech zohledněno, jelikož vzhledem k několika málo hodnotám, které přicházely v úvahu (18, 24, 30), by nebylo možné odhadnout systematickou chybu modelu.

V obou modelech byl však navíc vzat v úvahu zajímavý fenomén zřejmý z výchozí věkové struktury (obr. 1), že totiž počínaje věkem 24 je věk zakončený na 6 výrazně četnější (někde více než dvojnásobně) než věk zakončený na 4. To mohlo být způsobeno tím, že v případě věku 23, 33 atd. bylo zaokrouhlováno nahoru nikoli na 24, 34 atd., ale přímo na 25, 35 atd. a dále že věk 25, 35 atd. byl zaokrouhlován pouze nahoru na 26, 36 atd. a nikoli naopak.

Aby bylo možné odhadnout četnosti h_{10}, h_5, h_3, které v empirických hodnotách věku zakončených na 0, 5 a sudé číslo připadaly na zaokrouhlení, a odtud z empirických hodnot y_i odhadnout teoretické hodnoty Y_i, předpokládalo se v obou modelech, že se sousední teoretické hodnoty věku Y_{28}, Y_{29}, Y_{30}, Y_{31}, Y_{32} liší o stejnou hodnotu Δ; stejně tak hodnoty Y_{33}, Y_{34}, Y_{35}, Y_{36}, Y_{37}, atd.

4. Rovnoměrný model

V rovnoměrném modelu předpokládáme, že četnosti hodnot zaokrouhlených na věk zakončený 0, 5, a sudým číslem jsou pro všechny okolní zaokrouhlované hodnoty stejné. Např. při zaokrouhlování na věk 30 byl zaokrouhlen stejný počet $h_{10}(30)$ osob o skutečném věku 26, 27, 28, 29, 31, 32, 33, 34, při zaokrouhlování na věk 35 byl zaokrouhlen stejný počet $h_5(35)$ osob o skutečném věku 33,34,37 (k věku 36 viz předposlední odstavec kapitoly 3), atd.

Z rovnic

$$y_{28} = Y_{30} + 2\Delta - h_{10} + 2h_3(28) \tag{2}$$

$$y_{29} = Y_{30} + \Delta - h_{10} - h_3(28) \tag{3}$$

$$y_{30} = Y_{30} + 8h_{10} \tag{4}$$

$$y_{31} = Y_{30} - \Delta - h_{10} - h_3(32) \tag{5}$$

$$y_{32} = Y_{30} - 2\Delta - h_{10} + 2h_3(32) \tag{6}$$

vypočteme neznámé $Y_{30}, h_{10}, h_3(28), h_3(32), \Delta$ a zbývající hodnoty Y_{28}, Y_{29}, Y_{31}, Y_{32} a stejně postupujeme u obdobných rovnic pro $y_{38}, y_{39}, y_{40}, y_{41}, y_{42}$, atd.

Podobně vyřešíme soustavu

$$y_{33} = Y_{35} + 2\Delta - h_{10}(30) - h_5 - h_3(34) - h_3(32) \tag{7}$$

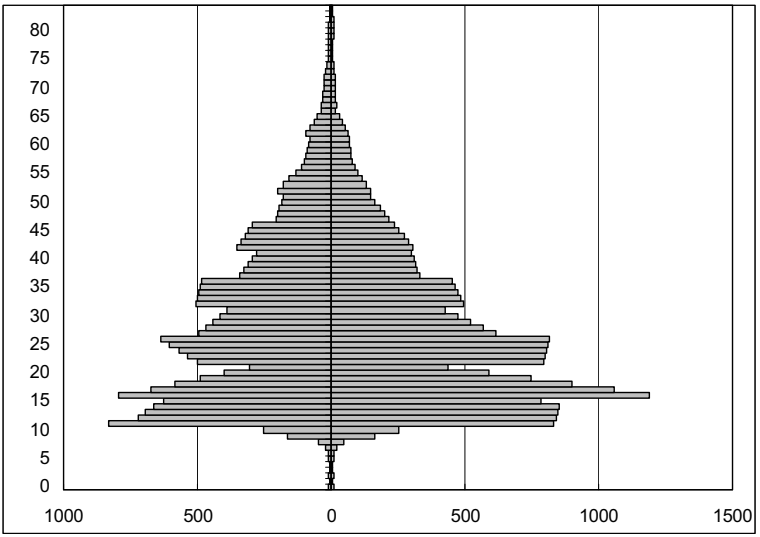
$$y_{34} = Y_{35} + \Delta - h_{10}(30) - h_5 + h_3(34) \tag{8}$$

$$y_{35} = Y_{35} + 4h_5 - h_3(36) \tag{9}$$

$$y_{36} = Y_{35} - \Delta - h_{10}(40) - h_5 + 2h_3(36) \tag{10}$$

$$y_{37} = Y_{35} - 2\Delta - h_{10}(40) - h_5 - h_3(36) - h_3(38) \tag{11}$$

a obdobné soustavy pro $y_{43}, y_{44}, y_{45}, y_{46}, y_{47}$ atd.



Obrázek 4: Rovnoměrný model

5. Lineární model

V lineárním modelu se předpokládá, že četnosti hodnot zaokrouhlených na věk zakončený 0, 5 a sudým číslem jsou tím větší, čím je zaokrouhlovaná hodnota blíže k té, na kterou se zaokrouhluje. Např. při zaokrouhlování na věk 30 by se zaokrouhlovalo $h_{10}(30)$ osob o skutečném věku 26 a 34, dvojnásobný počet osob $2h_{10}(30)$ o skutečném věku 27 a 33, $3h_{10}(30)$ osob o skutečném věku 28 a 32, a $4h_{10}(30)$ osob o skutečném věku 29 a 31, a podobně při zaokrouhlování na věk 35 (zde se však zohlednil vliv fenoménu 34, 36 – viz předposlední odstavec kapitoly 3), atd.

Podobně jako v rovnoměrném modelu vycházíme z rovnic

$$y_{28} = Y_{30} + 2\Delta - 3h_{10} + 2h_3(28) \tag{12}$$

$$y_{29} = Y_{30} + \Delta - 4h_{10} - h_3(28) \tag{13}$$

$$y_{30} = Y_{30} + 20h_{10} \tag{14}$$

$$y_{31} = Y_{30} - \Delta - 4h_{10} - h_3(32) \tag{15}$$

$$y_{32} = Y_{30} - 2\Delta - 3h_{10} + 2h_3(32) \tag{16}$$

a rovnic

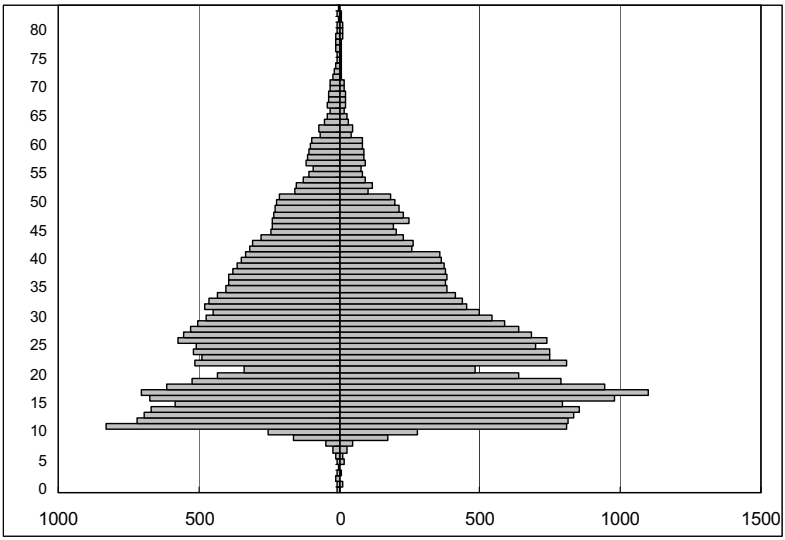
$$y_{33} = Y_{35} + 2\Delta - 2h_{10}(30) - h_5 - h_3(34) - h_3(32) \tag{17}$$

$$y_{34} = Y_{35} + \Delta - h_{10}(30) - 2h_5 + h_3(34) \tag{18}$$

$$y_{35} = Y_{35} + 4h_5 - h_3(36) \tag{19}$$

$$y_{36} = Y_{35} - \Delta - h_{10}(40) + 2h_3(36) \tag{20}$$

$$y_{37} = Y_{35} - 2\Delta - 2h_{10}(40) - h_5 - h_3(36) - h_3(38) \tag{21}$$



Obrázek 5: Lineární model

6. Závěr

Z obou teoretických bihistogramů (obr. 4 a 5) je patrné, že v případě studované populace vykazuje menší systematickou chybu (periodicitu teoretických dat) lineární model. V případě rovnoměrného modelu se zdá, že tento model věrněji vystihuje ženskou složku populace.

7. Literatura

[1] DUCHÁČEK, K. – FIALOVÁ, L. – HORSKÁ, P. – RÉPÁSOVÁ, M. – SLÁDEK, M.: On using the 1661–1839 lists of subject of the Třeboň dominion to study the age structure of the population, in: HD 13, 1989, s. 59–75.

[2] FIALOVÁ, L. – KUČERA, M. – MAUR, E. – HORSKÁ, P. – MUSIL, J. – STLOUKAL, M.: Dějiny obyvatelstva českých zemí, Praha, 1998.

[3] HENZLER, J. A KOLEKTIV: Matematika pro ekonomy, Oeconomica, Praha, 2007.

[4] KAČEROVÁ, E.: Struktura obyvatelstva Choceňska v roce 1651, Demografie [online], 2008, s. 1–4, http://www.demografie.cz

[5] MAUR, E.: Problémy demografické struktury Čech v polovině 17. století, in: ČsČH XIX, 1971, s. 849–850.

[6] PAZDEROVÁ, A.: Soupis poddaných podle víry – Chrudimsko, Národní archív, 2002.

[7] ŽVÁČEK, J. – HENZLER, J.: Statistika pro ekonomy, interaktivní text na CD, MUP, Praha, 2009.

Adresa autorů:

Eva Kačerová, RNDr.
katedra demografie FIS VŠE
nám. W. Churchilla 3
130 67 Praha 3
kacerova@vse.cz

Jiří Henzler, doc. RNDr. CSc.
katedra matematiky FIS VŠE
nám. W. Churchilla 3
130 67 Praha 3
henzler@vse.cz

Systém manažérstva kvality podľa požiadaviek normy ISO 9001 v ambulantných zdravotníckych zariadeniach
ISO 9001 Quality Management System in Outpatient Health Care Facilities

Kostičová Michaela, Knezović Renáta, Badalík Ladislav

Abstract: The partial results of the national questionnaire survey involving 671 respondents – outpatient health care providers are presented in this article. The aim of the survey was to analyse the correlation between education of outpatient health care providers in the area of quality management system (QMS) and their attitude towards the implementation of an ISO 9001 QMS in outpatient health care facilities. The respondents agreed with all contributions of ISO 9001 QMS presented in the questionnaire, but the significant higher rate of acceptance was expressed by respondents who were more intensively and longer educated in QMS. The results pointed out that for understanding the contributions of QMS the importance of education in QMS is significant and especially doctors should be educated.

Key words: Questionnaire survey, outpatient health care providers, quality management system, ISO 9001:2000 standard, education
Kľúčové slová: Dotazníkový prieskum, poskytovateľ ambulantnej zdravotnej starostlivosti, systém manažérstva kvality, norma ISO 9001:2000, vzdelávanie

1. Úvod

V súlade s §9 zákona č. 578/2004 Z.z. o poskytovateľoch zdravotnej starostlivosti (PZS), zdravotníckych pracovníkoch, stavovských organizáciách v zdravotníctve je povinnosťou každého PZS zabezpečovať systém kvality na dodržiavanie a zvyšovanie kvality tak, aby sa vzťahoval na všetky činnosti, ktoré môžu v zdravotníckom zariadení (ZZ) ovplyvniť zdravie osoby alebo priebeh jej liečby a aby personálne zabezpečenie a materiálno-technické vybavenie ZZ bolo v súlade s odborným zameraním ZZ a spĺňalo minimálne požiadavky ustanovené príslušným právnym predpisom. Systém kvality je písomne dokumentovaný systém, ktorého základným cieľom je znižovanie nedostatkov v poskytovaní zdravotnej starostlivosti pri súčasnom zvyšovaní spokojnosti pacientov a pri zachovaní ekonomickej efektívnosti poskytovateľa [9, 10].

Systém kvality v zdravotníctve môžeme definovať ako súhrn štruktúry organizácie, jednotlivých zodpovedností, procedúr, procesov a zdrojov, ktoré sú potrebné k trvalému zlepšovaniu kvality poskytovaných zdravotníckych služieb, ktorých konečným cieľom je zlepšovanie zdravotného stavu, zvyšovanie kvality života a spokojnosti obyvateľov, ktorým poskytujú starostlivosť. Systém kvality teda obsahuje celý proces tvorby postupov, zberu informácií, stanovenia štandardov a hodnotenia výsledkov toho, čo v zdravotníctve organizujeme ako zdravotnú starostlivosť zdravotnícke služby [5].

Legislatíva požaduje od PZS zabezpečovať a zdokumentovať systém kvality, ale nie sú stanovené príslušným právnym predpisom podrobnosti o zabezpečovaní systému kvality ani spôsob jeho kontroly. Jedine na báze dobrovoľnosti je možné získať oficiálne potvrdenie zavedenia SMK a to formou certifikácie ZZ podľa požiadaviek normy ISO 9001, prípadne akreditáciou medicínskych laboratórií Slovenskou národnou akreditačnou službou (SNAS). Nedostatočne využívaným v zdravotníctve je model excelentnosti EFQM (European Foundation for Quality Management - Európskej nadácie pre kvalitu) ako aj akreditácia ZZ medzinárodnou komisiou JCI (Joint Comission International). Pre malé ambulantné ZZ je získanie medzinárodne platného certifikátu kvality ISO 9001 v súčasnosti jedinou dostupnou

a pomerne často využívanou formou deklarovania kvality. Požiadavky ISO sú však skôr zamerané do oblasti výroby a používajú terminológiu líšiacu sa od tej, ktorá sa používa v oblasti zdravotníctva. Problematická je aplikácia ISO noriem na špecifiká zdravotníckych služieb. Nevýhodou sú aj finančné náklady a komercionalizácia certifikácie (množstvo poradenských a certifikačných spoločnosti), [2].

Zavedenie systému SMK postupne zasiahne ako klinickú tak ambulantnú zložku zdravotníctva. Vzhľadom ku všeobecnému pojmovému zmätku a nepriehľadnosti najrôznejších systémov zavádzania kvality je nutné vniesť do tejto oblasti viac svetla najmä sústredenými edukačnými aktivitami. Len pri užití najadekvátnejšej metódy pre danú špecifickú zdravotnú oblasť môžeme očakávať jednak jej prijatie, jednak jej ďalší rozvoj. Jedine týmto spôsobom je možné zabrániť účelovej komercionalizácií certifikácií a akreditácií.

2. Ciele

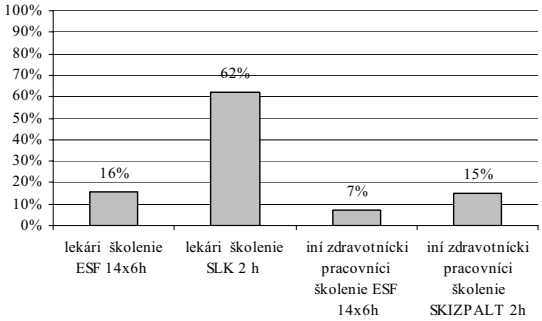
Jedným z cieľov dotazníkového prieskumu bolo analyzovať vzťah medzi vzdelávaním PAZS v oblasti SMK a ich vnímaním významu a prínosov implementácie SMK podľa požiadaviek normy ISO 9001 v zdravotníckych zariadeniach ambulantného typu. Východiskom k realizácii výskumu bol predpoklad nízkej informovanosti PAZS o prínosoch, význame a podstate SMK, ktorá spolu s nejasnými legislatívnymi požiadavkami, preťaženosťou zdravotníckych pracovníkov a nedostatkom ľudských a finančných zdrojoch v zdravotníctve môže v negatívnom smere ovplyvniť postoj PAZS k implementácii SMK v ZZ. Výsledky analýz môžu byť podkladom pre optimalizáciu vzdelávania v oblasti kvality a manažmentu u zdravotníckych pracovníkov a podnetom k legislatívnym úpravám, ktoré by jednoznačnejšie špecifikovali požiadavky na zabezpečovanie systému kvality a motivovali PZS k jeho implementácii.

3. Súbor a metódy

Dotazníkový prieskum bol realizovaný v období máj 2007 až apríl 2008 na vzorke 671 PAZS. Dotazník bol neštandardizovaný, anonymný, obsahoval zatvorené, polootvorené a otvorené položky a bol distribuovaný PAZS na konci vzdelávacej aktivity, ktorú absolvovali. Distribuovaných bolo 1000 dotazníkov návratnosť bola 67,1%. Súbor reprezentovalo 520 (78%) ambulantných lekárov a 151 (22%) iných zdravotníckych pracovníkov. 156 (23%) respondentov z bratislavského kraja absolvovalo v rámci projektu Európskeho sociálneho fondu (ESF) cyklus 14 školení v oblasti SMK, v rozsahu jedno školenie 6h. 515 (77%) respondentov zo všetkých krajov Slovenska absolvovalo jednorázové 2 hodinové školenie zamerané na stručné sprístupnenie problematiky SMK organizované príslušnou stavovskou komorou (Slovenská lekárska komora - SLK, Slovenská komora iných zdravotníckych pracovníkov, asistentov, laborantov a technikov - SKIZPALT). Najpočetnejšia skupinou respondentov 62% boli ambulantní lekári, ktorí absolvovali 2h školenie SLK, 16% súboru tvorili ambulantní lekári absolventi cyklu školení ESF, 15% iní zdravotnícki pracovníci účastníci 2h školenia SKIZPALT a len 7% respondentov predstavovali iní zdravotnícki pracovníci, ktorí absolvovali školenia ESF (tabuľka 1, obrázok 1).

Tabuľka 1: Štruktúra respondentov podľa povolania a typu absolvovaného školenia

súbor	počet	Validné percentá	Kumulatívne percentá
iní zdravotnícki pracovníci školenie ESF 14x6h	50	7,5	7,5
ambulantní lekáriškolenie ESF 14x6h	106	15,8	23,2
ambulantní lekári školenie SLK 1x2h	414	61,7	84,9
iní zdravotnícki pracovníci školenie SKIZPALT 1x2h	101	15,1	100,0
spolu	671	100,0	



Obrázok 1: Štruktúra respondentov podľa povolania a typu absolvovaného školenia

Na štatistickú analýzu údajov sme použili metódy deskriptívnej štatistiky (frekvenčné tabuľky), základné štatistické charakteristiky (aritmetický priemer, smerodajná odchýlka) a metódy induktívnej štatistiky (Studentov t-test, jednostupňovú analýzu rozptylu ANOVA). Na štatistickú analýzu dát bol použitý štatistický softvér SPSS, verzia 14.

4. Výsledky

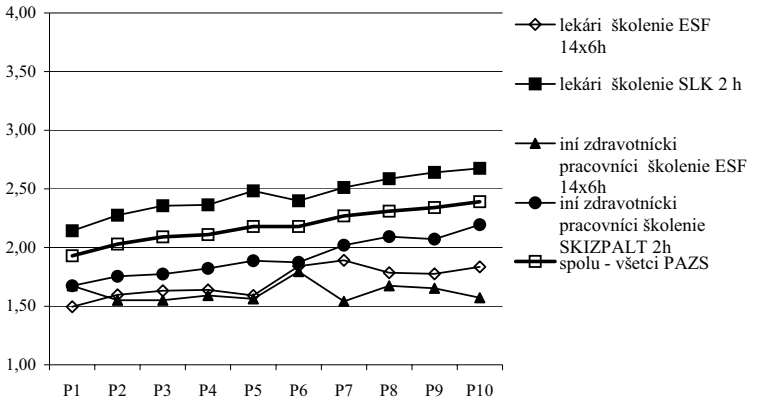
Respondenti mali v jednej časti dotazníka prezentovať do akej miery súhlasia s konkrétnymi 10 prezentovanými prínosmi (označené symbolmi P1-P10) implementácie SMK v súlade s požiadavky normy ISO 9001 pre činnosť ZZ ambulantného typu. Mieru súhlasu PAZS s jednotlivými prínosmi vyjadrili označením možnosti jednoznačne súhlasím, skôr súhlasím, skôr nesúhlasím, jednoznačne nesúhlasím. Pre porovnávanie jednotlivých podsúborov v miere súhlasu sme vytvorili intervalovú stupnicu v rozsahu 1-4, pričom hodnota 1 znamenala jednoznačne súhlasím, a hodnota 4 jednoznačne nesúhlasím. Za cely súbor respondentov ako aj za jednotlivé podsúbori sme vypočítali aritmetický priemer vyjadrujúci priemernú mieru súhlasu s konkrétnym prezentovaným prínosom implementácie systému kvality. Aritmetické priemery sme zoradili ako za celý súbor respondentov tak za jednotlivé podsúbori vzostupne, čím sme získali poradie jednotlivých prínosov z hľadiska ich preferencií respondentmi vzhľadom na typ absolvovaného školenia a povolanie. Rozdiely medzi jednotlivými podsúbormi sme testovali jednostupňovou analýzou rozptylu ANOVA. Celková priemerná miera súhlasu celého súboru a jednotlivých podsúborov respondentov so všetkými prezentovanými prínosmi nám poskytla údaj o tom, či PAZS súhlasia, že zavedený SMK je vo všeobecnosti prínosom pre ZZ ambulantného typu. Analýza reliability nám potvrdila vysokú mieru spoľahlivosti tejto časti dotazníka (Cronbachov alfa = 0,96).

Celková priemerná miera súhlasu s prínosmi SMK za celý súbor respondentov bola 2,18±0,80 čo na škále 1-4 znamená, že respondenti skôr súhlasia (priemer <2,5), že zavedenie SMK v ZZ ambulantného typu má svoje prínosy a väčšina z nich súhlasí aj s jednotlivými konkrétnymi prínosmi. Väčšina PAZS si myslí, že zavedenie systému kvality predovšetkým zvýši právnu ochranu PAZS (1,93±0,92), zabezpečí administratívny poriadok v ZZ (2,03±0,89), vytvorí jasné pravidlá fungovania ZZ (2,09±0,95) a pozdvihne jeho imidž (2,11±0,93). Respondenti zároveň súhlasia, že uplatňovanie požiadaviek normy ISO 9001 na SMK je príležitosťou zlepšiť organizáciu práce (2,18±0,93), objektívne hodnotiť výsledky ZZ (2,18±0,90) a motivovať k trvalému zlepšovaniu (P7, 65,2%, 2,27±0,92). V najmenšej miere respondenti súhlasili, že systém kvality zefektívni činnosti ZZ (2,31±0,95), zvýši

kvalitu poskytovaných služieb (2,34±0,97) a spokojnosť pacientov (2,39±1,02), v týchto položkách sme medzi jednotlivými podsúbormi zaznamenali štatisticky najvýznamnejšie rozdiely (tabuľka 2, obrázok 2). V miere súhlasu aj s ostatnými prínosmi boli medzi podsúbormi signifikantné rozdiely ($p<0,05$). Respondenti, ktorí absolvovali cyklus školení ESF vo väčšej miere súhlasia s prínosmi v porovnaní s absolventmi jednorázovej vzdelávacej aktivity. Štatisticky významné rozdiely ($p<0,05$) sme zaznamenali pri porovnávaní podsúborov ambulantných lekároch aj zdravotníckych pracovníkov.

Tabuľka 2: Miera súhlasu respondentov s prínosmi implementácie SMK podľa požiadaviek normy ISO 9001:2000

Súbor	Miera súhlasu (škála 1-4)	Prínosy implementácie SMK v súlade s požiadavkami normy ISO 9001									
		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
lekári školenie ESF 14x6h	počet	103	102	103	100	103	101	101	103	103	103
	aritm. priemer	1,5	1,6	1,63	1,64	1,59	1,84	1,89	1,79	1,78	1,83
	smer. och.	0,74	0,80	0,80	0,76	0,62	0,81	0,71	0,79	0,77	0,79
lekári školenie SLK 2 h	počet	398	397	394	396	398	390	395	392	400	399
	aritm. priemer	2,14	2,27	2,36	2,36	2,48	2,40	2,51	2,59	2,64	2,67
	smer. och.	0,96	0,91	0,98	0,96	0,92	0,90	0,95	0,95	0,96	1,04
	sig	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
iní zdravotnícki pracovníci školenie ESF 14x6h	počet	46	49	49	49	48	49	48	49	49	49
	aritm. priemer	1,67	1,55	1,55	1,59	1,56	1,8	1,54	1,67	1,65	1,57
	smer. och.	0,82	0,74	0,68	0,70	0,74	0,76	0,65	0,69	0,83	0,68
iní zdravotnícki pracovníci školenie SKIZPALT 2h	počet	98	98	97	96	98	95	98	97	98	98
	aritm. priemer	1,67	1,76	1,77	1,82	1,89	2,18	2,02	2,09	2,07	2,19
	smer. och.	0,70	0,63	0,69	0,67	0,70	0,90	0,70	0,74	0,71	0,77
	sig.	0,20	0,16	0,20	0,13	0,03	0,87	0,00	0,00	0,00	0,00
spolu	počet	645	646	643	641	647	635	642	641	650	649
	aritm. priemer	1,93	2,03	2,09	2,11	2,18	2,18	2,02	2,31	2,34	2,39
	smer. och.	0,92	0,90	0,95	0,93	0,93	0,90	0,70	0,95	0,97	1,02
	sig.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	aritm. priemer										2,18
	smer. och.										0,80



Obrázok 2.: Miera súhlasu respondentov s prínosmi implementácie SMK podľa požiadaviek normy ISO 9001:2000

Legenda

P1 Zvýšenie právnej ochrany PAZS
P2 Zabezpečenie administratívneho poriadku
P3 Vytvorenie jasných pravidiel fungovania ZZ
P4 Pozdvihnutie imidžu ZZ
P5 Zlepšenie organizácie práce
P6 Príležitosť objektívne hodnotiť výsledky ZZ
P7 Motivácia k trvalému zlepšovaniu
P8 Zefektívnenie činnosti/procesy
P9 Zvýšenie kvality poskytovaných služieb
P10 Zvýšenie spokojnosti pacientov

Ambulantní lekári (reprezentujúci 78% všetkých respondentov) bez ohľadu na rozsah absolvovanej vzdelávacej aktivity zdieľajú názor väčšiny PAZS a to, že uplatňovanie požiadaviek normy ISO 9001 na SMK v ZZ predovšetkým zvýši právnu ochranu PAZS (P1), zabezpečí administratívny poriadok v ZZ (P2), vytvorí jasné pravidlá fungovania ZZ (P3), pozdvihne jeho imidž (P4) a zlepší organizáciu práce (P5). Ambulantní lekári – účastníci jednorázového školenia ako jediná skupina respondentov vyjadrila názor, že zavedenie systému kvality nezvýši spokojnosť pacientov (2,67±1,02) ani kvalitu poskytovaných služieb (2,64±0,96), neprispeje k zefektívneniu činnosti ZZ (2,59±0,95) a nemotivuje ich to k zlepšovaniu (2,51±0,95). Tento podsúbor respondentov síce súhlasil s ostatnými prínosmi, ale v porovnaní s názormi ostatných respondentov v najmenšej miere. V porovnaní so skupinou lekárov, ktorí absolvovali cyklus školení ESF sme zaznamenali signifikantné rozdiely (p=0,000) v miere súhlasu so všetkými prezentovanými prínosmi, k rovnakému záveru sme dospeli aj pri porovnaní s ostatnými podsúbormi (p=0,000). Ambulantní lekári, ktorí absolvovali cyklus školení ESF súhlasia so všetkými prínosmi, v najmenšej miere súhlasia, že zavedený SMK umožňuje objektívne hodnotiť výsledky ZZ (1,84±0,81) a motivuje k trvalému zlepšovaniu (1,89±0,71), (tabuľka 2, obrázok 2).

Iní zdravotnícki pracovníci vyjadrili súhlas so všetkými prezentovanými prínosmi a to bez ohľadu na rozsah absolvovaného školenia. Signifikantné rozdiely boli v miere súhlasu len s niektorými prínosmi. Absolventi cyklu školení ESF v porovnaní s absolventmi jednorázového školenia v signifikantne väčšej miere súhlasia, že implementácia SMK predovšetkým motivuje PAZS k trvalému zlepšovaniu (1,54±0,65), zvýši spokojnosť pacientov (1,57±0,68,\), zefektívni procesy v ZZ (1,67±0,69), zlepší organizáciu práce (1,56±0,74), zvýši kvalitu poskytovaných služieb (1,65±0,830). Táto skupina respondentov najmenej zo všetkých prínosov súhlasí, že SMK umožní objektívne hodnotenie výsledkov ZZ (1,8±0,76) a zvýši ich právnu ochranu (1,67±0,82). Iní zdravotnícki pracovníci, ktorí absolvovali len jednorázové školenie rovnako v prevažnej miere súhlasia so všetkými prínosmi ale najmenej súhlasia s prínosmi systému kvalitu na efektivitu procesov (2,09±0,74), kvalitu služieb (2,07±0,71) a spokojnosť pacientov (2,19±0,77), (tabuľka 2, obrázok 2)

5. Diskusia

Na základe výsledkov môžeme konštatovať, že PAZS súhlasia s prínosmi zavedeného SMK podľa požiadaviek normy ISO 9001:2000 pre činnosť ZZ ambulantného typu V miere súhlasu s jednotlivými prínosmi sme však zaznamenali medzi respondentmi signifikantné rozdiely. Signifikantne vyššiu mieru súhlasu s týmito prínosmi deklarovali absolventi cyklu školení. Najmenšiu mieru súhlasu s jednotlivými prínosmi prezentoval podsúbor ambulantných lekárov – absolventov jednorázového školenia, ktorí si zároveň ako jediná skupina respondentov myslí, že zavedenie SMK nezvýši spokojnosť pacientov ani kvalitu poskytovaných služieb a neprispeje k zefektívneniu činnosti ZZ. Zdravotnícki pracovníci súhlasia so všetkými prínosmi bez ohľadu na rozsah absolvovaného školenia, vzdelávania však u nich prispieva k zvýšeniu miery súhlasu s niektorými prínosmi.

To, že ambulantní lekári vnímajú prínosy SMK potvrdil aj celoslovenský dotazníkový prieskum autorov z MTF STU v Trnave na vzorke 145 ambulantných lekárov, podľa ktorého len 15% lekárov si myslí, že SMK nezvýši kvalitu služieb, 30% to nevie posúdiť a 55% lekárov si myslí, že zavedenie SMK zvýši predovšetkým kvalitu ich odbornej práce, zavedie poriadok v administratíve a naplní legislatívnu požiadavku [8]. Na praktické prínosy zavedeného SMK u PAZS ako sú: preukázanie plnenia stanovených postupov a špecifikácií, efektivita vykonávaných činností, optimalizácia nákladov, získanie a potvrdenie dôvery pacientov, partnerov a všetkých zainteresovaných strán a získanie nových klientov poukazujú aj odborníci zo Slovenskej únie špecialistov [4].

6. Záver

Výsledky poukazujú na zložitosť problematiky SMK, ktorá si vyžaduje rozsiahle a intenzívne vzdelávanie, ktoré jednoznačne prispieva k pochopeniu významu a prínosov implementácie SMK aj pre ZZ ambulantného typu. Naliehavá je potreba vzdelávania najmä u lekárov a to ako na pregraduálnej tak postgraduálnej úrovni v oblasti zdravotníckeho manažmentu so zameraním aj na riadenie kvality [6].

Dôležité je, aby si PZS uvedomili, že implementácia SMK je zameraná predovšetkým na trvalé zlepšovanie kvality a to sa nekoná automaticky, vyžaduje si systematické riadenie zamerané na zlepšovanie starostlivosti, čo znamená nielen implementovanie špecifických zmien, ale aj transformáciu kultúry organizácie na kultúru, kde je každý angažovaný do procesu trvalého zlepšovania kvality a má na to aj vedomosti [1].

Neexistuje jediný najlepší prístup k zlepšovaniu kvality. Všetky modely systémov manažérstva kvality vychádzajú zo spoločného základu, postupne sa dostávajú z oblasti priemyslovej a služieb do sféry zdravotníctva. Pri správnom užití, rešpektovaní národných špecifík a špecifík jednotlivých druhov zdravotníckych zariadení vedú k zdokonaleniu a zladeniu tak zložitého mechanizmu, akým je poskytovanie zdravotnej starostlivosti [3].

Kvalitu nemožno nariadiť ani dosiahnuť plnením noriem, legislatívnych požiadaviek, dodržiavaním štandardov, tá musí vychádzať z vnútorného presvedčenia a angažovanosti ako vedenia tak všetkých pracovníkov zdravotníckeho zariadenia. Na otázku „Aké je tajomstvo kvality?“ odpovedal profesor A.Donabedian – otec kvality zdravotnej starostlivosti na sklonku svojho života: „Veľmi jednoduché, je to láska – láska k vedomostiam, láska k človeku a láska k Bohu“ [7].

7. Literatúra

[1] BAILY, M. – BOTRELL, M. – LYNN, J. – JENNINGS, B. 2006. The ethics fo using QI methods to improve health care quality and safety. A Hasting center special report, 2006. 34 s.

[2] BOUREK, A. A KOL. 2001-2004. Programy kvality a standardy léčebných postupů: Praktická příručka pro nemocnice, polikliniky a ambulantní péči. Praha: Verlag Dashöfer, 2001-2004. ISBN 80-86229-29-7.

[3] BUETOW, S. A. 2004. New Zeland Mori quality improvement in health care: lessons from an ideal type. In: International Journal for Quality in Health Care., č.5, 2004, s. 417-22.

[4] BUNGANIČ, I. 2005. Informácia k zavádzaniu systému manažérstva kvality v špecializovanej ambulantnej starostlivosti. In: Gastroenterológia pre prax, č.3, 2005, s.157-158.

[5] GLADKIJ, IVAN A KOL. 2003. Management ve zdravotnictví. Brno: Computer Press, 2003. 380s. ISBN 80-7226-996-8.

[6] KOSTIČOVÁ, M. – OZOROVSKÝ, V. 2008. Trendy vo výučbe zdravotníckeho manažmentu na Lekárskej fakulte UK v Bratislave. In: Současne trendy ve vzdělávání budoucích lékařů.

Sborník z konference. 1.vydání. Eds: Králová, J., Raková, M., Husárová, A. Olomouc: Univerzita Palackého v Olomouci, Lékařská fakulta, 2008, 79s. ISBN 978-80-244-2198-8.
[7] SUŇOL, R. 2000. Avedis Donabedian. In: International Journal for Quality in Health Care, č.6, 2000, s. 451-454.
[8] ŠALGOVIČOVÁ, J.2007. Manažérstvo kvality v zdravotníctve. Plánovanie kvality. Trnava: Tripsoft, Slovenská technická univerzita v Bratislave, Materiálovotechnologická fakulta v Trnave, 2007, 135s.
[9] Zákon č. 578/2004 Z.z. o poskytovateľoch zdravotnej starostlivosti, zdravotníckych pracovníkoch, stavovských organizáciách v zdravotníctve a o zmene a doplnení niektorých zákonov v znení neskorších predpisov.
[10] Zákon č. 653/2007 Z.z. ktorým sa mení zákon č. 578/2004 Z. z. o poskytovateľoch zdravotnej starostlivosti, zdravotníckych pracovníkoch, stavovských organizáciách v zdravotníctve a o zmene a doplnení niektorých zákonov v znení neskorších predpisov.

Adresa autorov:

Kostičová Michaela, MUDr.
Knezović Renáta, Mgr.
Badalík Ladislav, prof., MUDr., DrSc.
Lekárska fakulta UK v Bratislave
Ústav sociálneho lekárstva a lekárskej etiky
Sasinkova 2
813 72 Bratislava
michaela.kosticova@fmed.uniba.sk
renata.knezovic@fmed.uniba.sk

Chudoba na Slovensku
Poverty in Slovakia

Viera Labudová

Abstract: Poverty is a worldwide problem that afflicts even the wealthiest countries in the world. The main aim of this contribution is to compare the regions of Slovak Republic according to data about incomes, poverty and social exclusion of Slovakian households. To compare regions there were used method of arrangement of multi-characteristics objects.

Key words: Poverty, EU SILC, social exclusion, principal component analysis (PCA)

Kľúčové slová: Chudoba, EU SILC, sociálne vylúčenie, metóda hlavných komponentov

1. Úvod

Chudoba je v súčasnosti považovaná za vážny spoločenský problém, ktorý sa dotýka nielen krajín tzv. tretieho sveta, ale aj priemyselne rozvinutých krajín, nevynímajúc krajiny Európy. Európska komisia, v súvislosti s nevyhnutnosťou riešenia tohto problému, vyhlásila rok 2010 za Európsky rok boja proti chudobe a sociálnemu vylúčeniu.

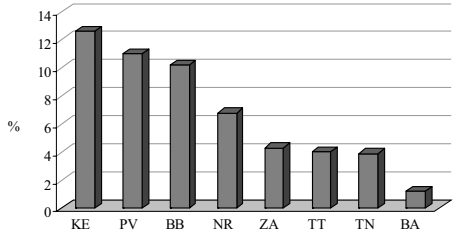
Cieľom tohto príspevku je porovnanie krajov Slovenskej republiky na základe ukazovateľov, ktoré sa používajú priamo pri meraní chudoby, alebo s problémom chudoby veľmi úzko súvisia.

Pre regionálne porovnanie sme sa rozhodli z toho dôvodu, že Slovensko je napriek malej rozlohe a ekonomickému potenciálu, krajinou s pomerne veľkými regionálnymi disparitami. Rozdiely medzi jednotlivými krajmi, ktoré sa prejavujú v dosiahnutej životnej úrovni, v stupni sociálno-ekonomického rozvoja, v ekonomickom a ľudskom potenciáli kraja a v možnostiach jeho ďalšieho rozvoja, sú výsledkom spolupôsobenia mnohých faktorov.

2. Meranie chudoby na Slovensku

Inštitúcie zaoberajúce sa meraním a analýzou chudoby používajú rôzne metódy skúmania chudoby. Opierajú sa o početné definície chudoby a vymedzenia hranice chudoby. Spoločným znakom týchto meraní je, že vychádzajú z ponímania chudoby ako stavu nedostatku materiálnych, sociálnych alebo kultúrnych zdrojov človeka, ktorý ho vylučuje z minimálne akceptovaného spôsobu života. Aj keď nie sú materiálne (peňažné) zdroje jedinou dimenziou chudoby, v dôsledku dostupnosti a relatívnej hodnovernosti údajov sa pri meraní chudoby najčastejšie používajú. [3]

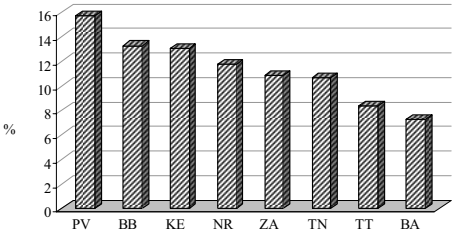
Za národný indikátor chudoby na Slovensku sa považuje životné minimum, alebo počet poberateľov a osôb závislých od dávky pomoci v hmotnej núdzi. Hranica životného minima nie je medzinárodne porovnateľným meradlom, ale je najdôležitejšia z politickej perspektívy, pretože odráža vládnu implicitnú definíciu chudoby. V roku 2006 bolo v systéme pomoci v hmotnej núdzi riešených priemerne mesačne 375 835 občanov, čo predstavovalo 7%-ný podiel na celkovom počte obyvateľov Slovenska. Najvyššie percento obyvateľstva v systéme pomoci v hmotnej núdzi za sledované obdobie bolo (obr. 1) v Košickom kraji (12,5%), Banskobystrickom kraji (10,2 %) a v Prešovskom kraji (11,0 %). Najnižší podiel poberateľov dávky mal Bratislavský kraj (1,2 %). [7]



Zdroj: Správa o sociálnej situácii obyvateľstva v roku 2006, vlastné spracovanie

Obrázok 1: Podiel obyvateľov v systéme pomoci v hmotnej núdzi na celkovom počte obyvateľov podľa krajov Slovenska

V Európskej únii sa pri charakterizovaní a porovnávaní chudoby medzi členskými štátmi kladie dôraz na indikátory relatívnej príjmovej chudoby meranej cez mieru rizika chudoby, ktorá je definovaná vo vzťahu k celkovému rozdeleniu príjmov. Miera rizika chudoby je definovaná ako podiel osôb s ekvivalentným disponibilným príjmom pod hranicou rizika chudoby, ktorou je 60% národného mediánu ekvivalentného príjmu. Podľa údajov ŠÚ SR získaných na základe harmonizovaného štatistického zisťovania EU SILC, miera rizika chudoby v roku 2006 na Slovensku dosiahla 11,6%. Miera rizika chudoby bola v roku 2006 najvyššia v Prešovskom (15,71%) a Banskobystrickom kraji (13,27%) a najnižšia v Bratislavskom kraji (7,29%).



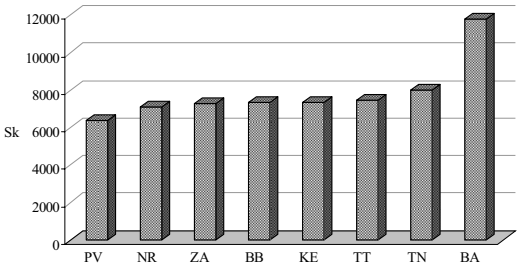
Zdroj: EU SILC 2006, ŠÚ SR, 2007, vlastné spracovanie

Obrázok 2: Podiel osôb pod hranicou chudoby (60% národného mediánu) v jednotlivých krajoch SR

Použitie týchto dvoch indikátorov chudoby pri hodnotení situácie v krajoch Slovenskej republiky prináša rozdielne výsledky. Usporiadanie krajov Slovenska sa mení vzhľadom na charakter použitej miery. Výnimkou je len Bratislavský kraj, ktorý použitím obidvoch indikátorov chudoby nadobudol najlepšie umiestnenie. Životná úroveň obyvateľov tohto kraja, ak ju posudzujeme na základe výšky priemerného disponibilného príjmu, je najvyššia.

Usudzujeme, že najnepriaznivejšia situácia je v Prešovskom kraji, ktorý patrí k trojici krajov (spolu s Košickým a Banskobystrickým) s najvyšším podielom obyvateľov v systéme pomoci v hmotnej núdzi, v ktorom sa pod hranicu chudoby ocitlo 15,71 % obyvateľov.

V Prešovskom kraji bola v roku 2006 najnižšia úroveň priemerného disponibilného príjmu domácností (6395 Sk/mesiac/1 osobu).



Zdroj: EU SILC 2006, ŠÚ SR, 2007, vlastné spracovanie

Obrázok 3: Priemerný disponibilný príjem domácností SR podľa krajov v roku 2006 (v Sk/mesiac/osobu)

3. Použitie viacrozmerných štatistických metód pri analýze regionálnych rozdielov

Pri porovnaní sociálnej situácie v jednotlivých krajoch Slovenska v roku 2006[1] sme použili nasledovné premenné:

MRCHU – *miera rizika chudoby* v % (podiel osôb s ekvivalentným disponibilným príjmom pod hranicou rizika chudoby, ktorou je 60% národného mediánu ekvivalentného príjmu)
Zdroj: EU SILC 2006, ŠÚ SR, 2007

MNE – nezamestnanosť v % (miera nezamestnanosti podľa výberového zisťovania pracovných síl (VZPS) k 31. 12. 2006)
Zdroj: Regdat, ŠÚ SR

DNE – *dlhodobá nezamestnanosť* v % (podiel nezamestnaných nad 12 mesiacov z celkového počtu evidovaných nezamestnaných k 31. 12. 2006)
Zdroj: Regdat, ŠÚ SR

VZDU – *vzdelanostná úroveň* v % (podiel ekonomicky aktívneho obyvateľstva vo veku 15+ s ukončeným základným vzdelaním alebo bez vzdelania na ekonomicky aktívnom obyvateľstve vo veku 15+ daného kraja)
Zdroj: Regdat, ŠÚ SR

SCNBY – *schopnosť splácať celkové náklady na bývanie* v % (podiel domácností, ktoré sa vyrovnávajú s platením zvyčajných výdavkov[2] s veľkými ťažkosťami z celkového počtu domácností kraja)
Zdroj: EU SILC 2006, ŠÚ SR, 2007

HMN – *obyvatelia v systéme pomoci v hmotnej núdzi* v % (podiel obyvateľov v systéme pomoci v hmotnej núdzi na celkovom počte obyvateľov)
Zdroj: Správa o sociálnej situácii obyvateľstva v roku 2006, MPSVaR, 2007

[1] Časová dimenzia analýzy bola ohraničená sprístupnením výsledkov štatistického zisťovania EU SILC.
[2] Pod zvyčajnými výdavkami sa rozumeli výdavky domácnosti na stravu, nájomné, úhradu za elektrinu, plyn, atď.[1]

MDP – *mesačný disponibilný príjem* v Sk (priemerný disponibilný príjem domácnosti prepočítaný na osobu a na mesiac)
Zdroj: EU SILC 2006, ŠÚ SR, 2007

Vzhľadom na to, že medzi dvojicami vybraných premenných existuje štatisticky významná závislosť (tabuľka 1), na porovnanie jednotlivých krajov sme použili umelé premenné , ktoré boli určené metódou hlavných komponentov.

Tabuľka 1: Hodnoty Pearsonovho koeficienta korelácie pre dvojice do analýzy vstupujúcich premenných

Pearson Correlation Coefficients, N = 8 Prob > |r| under H0: Rho=0							
	MRCHU	**CDP**	**MNE**	**DNE**	**NVZDU**	**HN**	**SCNBY**
MRCHU	1.00000	-0.73762 0.0367	0.88357 0.0036	0.87312 0.0046	0.63169 0.0929	0.88849 0.0032	0.77379 0.0243
CDP	-0.73762 0.0367	1.00000	-0.69313 0.0566	-0.86131 0.0060	-0.39991 0.3263	-0.64618 0.0834	-0.27804 0.5049
MNE	0.88357 0.0036	-0.69313 0.0566	1.00000	0.95183 0.0003	0.70008 0.0532	0.96938 <.0001	0.73864 0.0363
DNE	0.87312 0.0046	-0.86131 0.0060	0.95183 0.0003	1.00000	0.61975 0.1012	0.90853 0.0018	0.55772 0.1509
NVZDU	0.63169 0.0929	-0.39991 0.3263	0.70008 0.0532	0.61975 0.1012	1.00000	0.65169 0.0800	0.51602 0.1905
HN	0.88849 0.0032	-0.64618 0.0834	0.96938 <.0001	0.90853 0.0018	0.65169 0.0800	1.00000	0.73431 0.0380
SCNBY	0.77379 0.0243	-0.27804 0.5049	0.73864 0.0363	0.55772 0.1509	0.51602 0.1905	0.73431 0.0380	1.00000

Úlohu nájsť hlavné komponenty, ktorými môžeme nahradiť vstupnú množinu premenných sme riešili s využitím systému *SAS Enterprise Guid 5.2*.

V tabuľke 2 uvádzame vlastné čísla premenných pri jednotlivých hlavných komponentoch a podiel variability vysvetlenej príslušným komponentom na celkovej variabilite. Na usporiadanie jednotlivých krajov sme použili hodnoty prvého hlavného komponentu, ktorý vysvetľuje spolu 76,64 % variability vstupných premenných

Na základe hodnôt Pearsonovho koeficienta korelácie medzi pôvodnými premennými a prvým hlavným komponentom môžeme konštatovať silný vplyv všetkých premenných na hodnoty prvého hlavného komponentu.

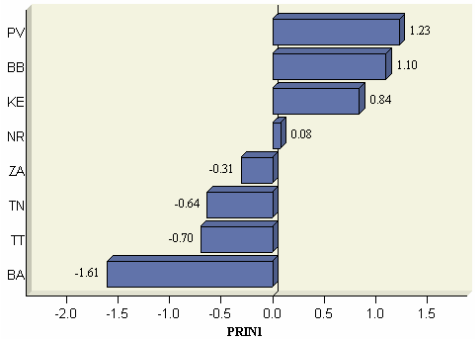
Tabuľka 2: Hodnoty vlastných čísel a variability vysvetlenej jednotlivými komponentmi.

Eigenvalues of the Correlation Matrix				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	5.36508187	4.54243046	0.7664	**0.7664**
2	0.82265140	0.31458292	0.1175	**0.8840**
3	0.50806848	0.30224043	0.0726	0.9565
4	0.20582805	0.13482603	0.0294	0.9859
5	0.07100202	0.04450216	0.0101	0.9961
6	0.02649986	0.02563155	0.0038	0.9999
7	0.00086831		0.0001	1.0000

Tabuľka 3: Hodnoty Pearsonovho koeficienta korelácie medzi premennými vstupujúcimi do analýzy a prvým hlavným komponentom

Pearson Correlation Coefficients, N = 8 Prob > |r| under H0: Rho=0							
	MRCHU	**MNE**	**DNE**	**NVZDU**	**HN**	**PRIN1**	**PRIN2**
PRIN1	0.95115 0.0003	0.97824 <.0001	0.95284 0.0003	0.73134 0.0392	0.95744 0.0002	1.00000	0.00000 1.0000

Pri usporiadaní krajov na základe hodnôt novovytvorenej premennej sme vychádzali z konštatovania negatívneho vplyvu[3] pôvodných premenných na situáciu v jednotlivých krajoch. To sa prejavuje aj v hodnotách komponentného skóre. Jeho kladné vysoké hodnoty indikujú regióny s relatívne vysokým výskytom chudoby (obrázok 4).



Obrázok 4: Poradie krajov SR určené na základe hodnôt komponentného skóre

[3] Podľa charakteru vplyvu na sledovaný jav sú tieto premenné tzv. destimulanty. Destimulanty dokazujú vyššiu úroveň rozvoja sledovaného javu svojimi nízkymi hodnotami

4. Záver

Cieľom tohto príspevku bolo porovnanie krajov Slovenskej republiky na základe charakteristík, ktoré sú silne späté s takým sociálnym fenoménom, ako je výskyt chudoby. Porovnanie krajov sme realizovali s využitím premenných, ktoré sa považujú za silné determinanty chudoby, prierezového indikátora chudoby a premennej vyjadrujúcej ekonomickú depriváciu domácností.

Najhoršia situácia je v Prešovskom, Banskobystrickom a Košickom kraji. Bratislavský kraj možno považovať za región, ktorý sa vzhľadom na sociálnu situáciu obyvateľov umiestnil na najlepšej pozícii. Priemerný disponibilný príjem tohto kraja je viac ako 1,8 krát vyšší ako v Prešovskom kraji, pričom podiel obyvateľov pod hranicou chudoby je v tomto kraji viac ako 2 krát nižší než v Prešovskom kraji.

5. Literatúra

[1] EU SILC 2006: Zisťovanie o príjmoch a životných podmienkach domácností v SR. 2007. Bratislava: ŠÚ SR, 2007.

[2] LIĎÁK, J.: Medzinárodná politika a problémy chudoby.(www.parlamentnykurier.sk/kur02-03/51.pdf)

[3] LIMUNKOVÁ, K.- VAGAČ, L.: Chudoba a regionálne rozdiely na Slovensku. (www.sdf.sk/sdf_media/9chudoba.pdf)

[4] Národný akčný plán sociálnej inklúzie 2004 – 2006. (www.employment.gov.sk/new/index.php?id=580)

[5] Slovenská republika: štúdia o životnej úrovni, zamestnanosti a trhu práce. 2001. Bratislava: Slovenská spoločnosť pre zahraničnú politiku, 2001. ISBN 80-968155-4-7.

[6] SONNTAGOVÁ, I., PILÁT, M.: Sociální politika v Evropské unii.(www.europeum.org)

[7] Správa o sociálnej situácii 2007. (http://www.employment.gov.sk/new/index.php?id=10264)

[8] STANKOVIČOVÁ, I.- VOJTKOVÁ, M. 2007. Viacrozmerné štatistické metódy s aplikáciami. Iura Edition, 2007. 261 s. ISBN 978-80-8078-152-1

Adresa autora:

Viera Labudová, RNDr. PhD.
Katedra štatistiky, FHI, Ekonomická univerzita
Dolnozemská 1/b, 852 35 Bratislava
e-mail: labudova@euba.sk

Oznam: Tento článok vznikol v rámci riešenia projektu VEGA č. **14586/07-202** *Modelovanie sociálnej situácie obyvateľstva a domácností v SR a jej regionálne a medzinárodné porovnania.*

Vybrané metódy štatistiky v aplikácii na teplotné pomery Nitry
Chosen Statistical Methods in Application to Temperature Conditions in Nitra

Peter Lenčéš, Janka Melušová

Abstract: This article deals trough chosen methods of mathematical statistics with actual problem of global warming. We analyse data about air temperature in area of Nitra in period of years 1995 – 2008 and compare them to the data from years 1961 - 1990. This period is considered as standard normal period. Our aim is to present the utility of statistics in the field of verification of potential changes in area of Nitra in years 1995 – 2008 in context of global warming.

Key words: Mathematical statistics, air temperature, statistical analysis, global warming

Kľúčové slová: Matematická štatistika, teplota vzduchu, štatistická analýza, globálne otepľovanie

1. Úvod

V ostatných rokoch sa čoraz častejšie začala dostávať do povedomia verejnosti problematika globálneho otepľovania. Na rôznych úrovniach totiž o nej diskutujú vedci, politici, novinári i bežní ľudia. Zdá sa, že diskusia o danej problematike je opodstatnená, keďže počnúc 70. rokmi minulého storočia a najmä v aktuálnom desaťročí odborníci zaznamenávajú trend vzostupu priemernej teploty vzduchu na Zemi, čo podľa [1] korešponduje s obsahom pojmu globálne otepľovanie.

Metódami matematickej štatistiky aplikovanými na údaje o teplote vzduchu z meteorologickej stanice Nitra – Veľké Janíkovce za obdobie rokov 1995 – 2008 sme sa pokúsili preukázať, či v tejto lokalite došlo v uvedenom období k prípadným zmenám v teplotných pomeroch v porovnaní s obdobím rokov 1961 – 1990, ktoré meteorológovia a klimatológia považujú za tzv. štandardné normálové obdobie, t. j. za určujúce v zmysle komparácie s aktuálnymi údajmi o meteorologických prvkoch.

2. Štatistická analýza

Meteorológovia a klimatológovia využívajú empirické údaje zo štandardného normálového obdobia na výpočet dlhodobých priemerných hodnôt teploty vzduchu – tzv. klimatických normálov – a ich komparáciou s aktuálnymi údajmi o teplote vzduchu získavajú prehľad o prípadných aberáciách. Tabuľka 1 obsahuje mesačné klimatické normály pre oblasť Nitry.

Tabuľka 1: Mesačné klimatické normály (Nitra – Veľké Janíkovce)

	mesiac											
	I.	II.	III.	IV.	V.	VI.	VII.	VIII.	IX.	X.	XI.	XII.
normál	-2,0	0,7	4,9	10,1	15,0	17,9	19,5	19,0	15,2	10,1	4,4	0,0

Zdroj: SHMÚ

V tabuľke 2 uvádzame hodnoty priemernej mesačnej teploty vzduchu pre oblasť Nitry vypočítané z údajov nameraných v období rokov 1995 – 2008.

Tabuľka 2: Priem. mesačná teplota vzduchu (Nitra – Veľké Janíkovce)

	mesiac											
rok	I.	II.	III.	IV.	V.	VI.	VII.	VIII.	IX.	X.	XI.	XII.
1995	-1,1	4,6	4,4	10,7	14,6	17,7	22,9	19,8	14,2	11,0	2,4	0,3
1996	-2,2	-3,1	1,9	11,0	16,5	19,3	18,3	19,4	11,9	10,5	7,0	-2,3
1997	-2,6	1,7	4,5	7,5	16,0	18,9	19,0	20,8	15,3	7,3	5,2	2,5
1998	2,0	4,5	4,1	12,1	15,4	19,8	21,0	20,9	15,2	10,8	1,9	-2,4
1999	-0,3	0,2	7,2	12,1	15,8	18,5	20,9	18,7	18,2	9,9	3,6	-0,3
2000	-2,4	2,5	5,0	13,9	17,2	20,4	18,8	21,6	15,1	13,3	8,2	2,2
2001	0,7	2,2	6,5	9,8	17,0	17,1	20,8	21,9	13,6	12,5	3,0	-6,0
2002	-1,4	4,0	6,6	10,3	18,0	20,0	22,6	21,0	14,9	9,1	7,8	-1,0
2003	-2,2	-2,0	5,1	10,2	17,9	21,9	21,7	22,8	15,7	7,8	6,7	1,1
2004	-3,1	1,1	4,6	11,7	14,1	18,1	20,3	20,6	15,5	12,1	5,6	0,9
2005	0,0	-2,4	3,0	11,3	15,5	18,4	20,8	18,9	16,9	11,0	4,0	0,3
2006	-3,8	-1,8	3,2	12,1	15,0	19,6	23,5	17,8	17,4	12,1	7,5	3,0
2007	4,1	4,6	7,6	11,9	16,9	20,8	22,2	21,6	13,5	9,3	3,5	-0,6
2008	1,7	2,6	5,6	11,3	16,3	20,6	20,6	20,1	15,0	11,1	6,9	2,9

Zdroj: SHMÚ

Do tabuľky 3 sme zaznamenali vypočítané odchýlky ($\Delta t_{pr.m.}$) hodnôt priemernej mesačnej teploty vzduchu z obdobia rokov 1995 – 2008 (tabuľka 2) oproti klimatickému normálu (tabuľka 1) a početnosti kladných a záporných odchýlok, t. j. početnosti výskytu teplotne nadpriemerných mesiacov ($\sum+$) a početnosti výskytu teplotne podpriemerných mesiacov ($\sum-$) prislúchajúce k jednotlivým rokom 1995 – 2008.

Tabuľka 3: Zmeny priem. mesač. teploty vzduchu voči normálu a početnosti klad. a záp. zmien

	mesiac													
	I.	II.	III.	IV.	V.	VI.	VII.	VIII.	IX.	X.	XI.	XII.		
rok	$\Delta t_{pr.m.}$												$\sum+$	$\sum-$
1995	0,9	3,9	-0,5	0,6	-0,4	-0,2	3,4	0,8	-1,0	0,9	-2,0	0,3	7	5
1996	-0,2	-3,8	-3,0	0,9	1,5	1,4	-1,2	0,4	-3,3	0,4	2,6	-2,3	6	6
1997	-0,6	1,0	-0,4	-2,6	1,0	1,0	-0,5	1,8	0,1	-2,8	0,8	2,5	7	5
1998	4,0	3,8	-0,8	2,0	0,4	1,9	1,5	1,9	0,0	0,7	-2,5	-2,4	8	3
1999	1,7	-0,5	2,3	2,0	0,8	0,6	1,4	-0,3	3,0	-0,2	-0,8	-0,3	7	5
2000	-0,4	1,8	0,1	3,8	2,2	2,5	-0,7	2,6	-0,1	3,2	3,8	2,2	9	3
2001	2,7	1,5	1,6	-0,3	2,0	-0,8	1,3	2,9	-1,6	2,4	-1,4	-6,0	7	5
2002	0,6	3,3	1,7	0,2	3,0	2,1	3,1	2,0	-0,3	-1,0	3,4	-1,0	9	3
2003	-0,2	-2,7	0,2	0,1	2,9	4,0	2,2	3,8	0,5	-2,3	2,3	1,1	9	3
2004	-1,1	0,4	-0,3	1,6	-0,9	0,2	0,8	1,6	0,3	2,0	1,2	0,9	9	3
2005	2,0	-3,1	-1,9	1,2	0,5	0,5	1,3	-0,1	1,7	0,9	-0,4	0,3	8	4
2006	-1,8	-2,5	-1,7	2,0	0,0	1,7	4,0	-1,2	2,2	2,0	3,1	3,0	8	4
2007	6,1	3,9	2,7	1,8	1,9	2,9	2,7	2,6	-1,7	-0,8	-0,9	-0,6	8	4
2008	3,7	1,9	0,7	1,2	1,3	2,7	1,1	1,1	-0,2	1,0	2,5	2,9	11	1

Zdroj: vlastný výpočet

Na základe údajov v tabuľke 3 môžeme vyvodiť, že pomer počtu teplotne nadpriemerných mesiacov k počtu teplotne podpriemerným mesiacom je 113 : 54, čo v nás evokuje, že obdobie rokov 1995 – 2008 bolo teplejšie, než štandardné normálové obdobie. Zo štatistického hľadiska však musíme byť pri prijatí hypotézy o tom, že obdobie rokov 1995 – 2008 bolo ako celok v porovnaní so štandardným normálovým obdobím teplejšie, opatrní. Totiž, chladnejšie mesiace sledovaného obdobia rokov 1995 – 2008 mohli byť teplotne podpriemerné až do takej miery, že takto kompenzovali ich nižší počet voči vyššiemu počtu teplotne nadpriemerných mesiacov a tým teplotné pomery oboch období vyrovnali. Hypotézu, že obdobie rokov 1995 – 2008 možno ako celok považovať za teplejšie než obdobie rokov 1961 – 1990, preveríme neskôr a to využitím štatistického testovania hodnôt priemernej ročnej teploty. Predtým sa však ešte bližšie zamerajme na početnosti výskytu teplotne nadpriemerných mesiacov (∑+) prislúchajúce k jednotlivým rokom (tabuľka 3). Tieto čísla nielenže prevyšujú početnosti výskytu teplotne podpriemerných mesiacov (∑–) prislúchajúce k jednotlivým rokom, ale – čo je dôležité – javí sa, že majú rastúci trend. To by však znamenalo, že otepľovanie v predmetnom období bolo vzhľadom na čas postupné, čo by korešpondovalo s povahou globálneho otepľovania. Využitím regresnej analýzy podľa [2] v MS EXCEL zistíme, či javiaci sa trend rastu početností teplotne nadpriemerných mesiacov možno považovať za štatisticky významný.

Tabuľka 4: Výstup regresnej analýzy v MS EXCEL

VÝSLEDEK

Regresní statistika	
Násobné R	0,688387724
Hodnota spolehlivosti R	0,473877658
Nastavená hodnota spoleh. R	0,43003413
Chyba stř. hodnoty	0,957905224
Pozorování	14

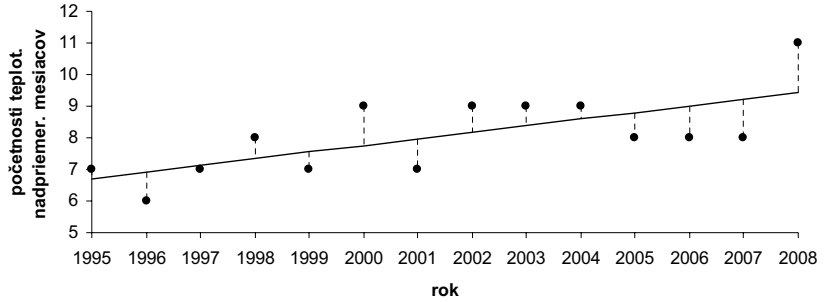
ANOVA

	Rozdíl	*SS*	*MS*	*F*	*Významnost F*
Regrese	1	9,917582418	9,917582418	10,80838323	0,006487977
Rezidua	12	11,01098901	0,917582418		
Celkem	13	20,92857143			

	Koeficienty	*Chyba stř. hodnoty*	*t stat*	*Hodnota P*	*Dolní 95%*
Hranice	-409,8241758	127,1125164	-3,224105599	0,007298677	-686,778557
Soubor X 1	0,208791209	0,063508498	3,287610566	0,006487977	0,070418079

Horní 95%	*Dolní 95,0%*	*Horní 95,0%*
-132,8697946	-686,778557	-132,8697946
0,347164339	0,070418079	0,347164339

Odhadli sme lineárny regresný model tvaru $Y = \beta_0 + \beta_1 x + e$. Jeho odhadom je regresná priamka s rovnicou $y = b_0 + b_1 x$. Koeficienty b_0 a b_1 nájdeme v ANOVA tabuľke v stĺpci *Koeficienty*. Pre tzv. lokujúcu konštantu platí: b_0 = –409,8241758. Tiež vidieť, že regresný koeficient je b_1 = 0,208791209. Regresný koeficient je smernicou regresnej priamky, ktorou sme v zmysle metódy najmenších štvorcov na obrázku 1 vystihli početnosti výskytu teplotne nadpriemerných mesiacov.



Obrázok 1: Regresná priamka s rovnicou y = – 409,8241758 + 0,208791209x

Podľa sklonu regresnej priamky (obrázok 1) je zrejmý trend rastu početností výskytu teplotne nadpriemerných mesiacov. Či je tento trend rastu štatisticky významný, preukážeme testom. Testujeme tzv. nulovú hypotézu H_0: $b_1 = 0$, podľa ktorej priamka s rovnicou $y = b_0 + b_1x$ má nulovú smernicu, proti alternatívnej hypotéze H_1: $b_1 \neq 0$, ktorá je negáciou nulovej hypotézy. Stačí len porovnať tzv. *Hodnotu P*, ktorú nájdeme v ANOVA tabuľke v riadku *Soubor X1* s hladinou významnosti napr. $\alpha = 0{,}05$. Keďže 0,006487977 < 0,05, nulovú hypotézu H_0 zamietame a prijímame hypotézu H_1. Teda smernicu b_1 = 0,208791209 regresnej priamky nemožno považovať za rovnú 0, čím sme preukázali, že trend rastu početností výskytu teplotne nadpriemerných mesiacov v období rokov 1995 – 2008 je štatisticky významný.

Teraz sa zameriame na hypotézu, podľa ktorej obdobie rokov 1995 – 2008 možno ako celok považovať za teplejšie než obdobie rokov 1961 – 1990. Využijeme tzv. dvojvýberový *t*-test s rovnosťou rozptylov. Budeme ním testovať rovnosť stredných hodnôt priemerných ročných teplôt vzduchu oboch období. Hodnoty priemernej ročnej teploty vzduchu pre oblasť Nitry vypočítané z údajov nameraných v štandardnom normálovom období nájdeme v tabuľke 5.

Tabuľka 5: Priem. ročná teplota vzduchu (Nitra – Veľké Janíkovce)

rok	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
$t_{pr.r.}$	10,3	8,7	8,7	9,0	8,2	10,0	10,1	9,7	9,1	9,2
rok	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
$t_{pr.r.}$	9,8	9,8	9,7	10,4	10,4	9,7	10,0	8,9	9,6	8,2
rok	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
$t_{pr.r.}$	9,9	10,0	10,2	9,2	8,5	9,3	9,0	10,0	10,5	10,6

Zdroj: SHMÚ

Obsahom tabuľky 6 sú hodnoty priemernej ročnej teploty vzduchu pre oblasť Nitry vypočítané z údajov nameraných v období rokov 1995 – 2008.

Tabuľka 6: Priem. ročná teplota vzduchu (Nitra – Veľké Janíkovce)

rok	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
$t_{pr.r.}$	10,1	9,0	9,7	10,5	10,4	11,3	9,9	11,0	10,6	10,1	9,8	10,5	11,3	11,2

Zdroj: SHMÚ

Skôr, ako na hodnoty priemernej ročnej teploty vzduchu oboch období aplikujeme dvojvýberový *t*-test s rovnosťou rozptylov, musíme využitím *F*-testu overiť, či rozptyly σ_1^2 a σ_2^2 hodnôt priemernej ročnej teploty vzduchu môžeme pre obe časové obdobia považovať za za zhodné. Budeme teda testovať nulovú hypotézu H_0: $\sigma_1^2 = \sigma_2^2$, ktorá hovorí o rovnosti rozptylov, proti alternatívnej hypotéze H_1: $\sigma_1^2 \neq \sigma_2^2$, ktorá tvrdí opak. V MS EXCEL podľa [2] zvolíme *Dvouvýběrový F-test pro rozptyl*. Po zadaní vstupných údajov, ktorými sú hodnoty priemernej ročnej teploty vzduchu pre obe obdobia, dostávame tabuľku 7.

Tabuľka 7: Výstup F-testu v MS EXCEL

Dvouvýběrový F-test pro rozptyl		
	Soubor 1	*Soubor 2*
Stř. hodnota	9,556667	10,38571
Rozptyl	0,468747	0,455165
Pozorování	30	14
Rozdíl	29	13
F	1,02984	
P(F<=f) (1)	0,499598	
F krit (1)	2,385906	

Zvolili sme hladinu významnosti α = 0,05. Dvojstrannú hodnotu *P* (riadok *P(F<=f) (1)* v tabuľke 7), t. j. 2 . 0,499598 = 0,999196, porovnáme s hladinou významnosti α = 0,05. Pretože 0,999196 > 0,05, hypotézu H_0 o rovnosti rozptylov hodnôt priemernej ročnej teploty vzduchu pre obe časové obdobia nezamietame. Využitím dvojvýberového *t*-testu s rovnosťou rozptylov možno teraz testovať rovnosť stredných hodnôt priemerných ročných teplôt vzduchu oboch období. Stanovíme nulovú hypotézu H_0: $\mu_1 = \mu_2$, podľa ktorej sa stredné hodnoty priemerných ročných teplôt oboch období rovnajú, proti alternatívnej hypotéze H_1: $\mu_1 \neq \mu_2$. V MS EXCEL uskutočníme tento raz *Dvouvýběrový t-test s rovností rozptylů*, pozri [2]. Tabuľku 8 sme dostali po zadaní vstupných údajov, ktorými sú hodnoty priemernej ročnej teploty vzduchu pre obe obdobia.

Tabuľka 8: Výstup dvojvýberového t-testu s rovnosťou rozptylov v MS EXCEL

Dvouvýběrový t-test s rovností rozptylů		
	Soubor 1	*Soubor 2*
Stř. hodnota	9,556666667	10,38571429
Rozptyl	0,468747126	0,455164835
Pozorování	30	14
Společný rozptyl	0,464543084	
Hyp. rozdíl stř. hodnot	0	
Rozdíl	42	
t stat	-3,7580693	
P(T<=t) (1)	0,000261154	
t krit (1)	1,681952358	
P(T<=t) (2)	0,000522307	
t krit (2)	2,018081679	

Dvojstrannú hodnotu *P* (riadok *P(T<=t) (2)* v tabuľke 8) 0,000522307 porovnáme s hladinou významnosti $\alpha = 0{,}05$. Pretože 0,000522307 < 0,05, hypotézu H_0 zamietame a prijímame hypotézu H_1, ktorá tvrdí, že stredné hodnoty priemerných ročných teplôt vzduchu štandardného normálového obdobia a obdobia rokov 1995 – 2008 sa významne odlišujú. Navyše, ako máme možnosť vidieť v tabuľke 8 (riadok *Stř. hodnota*), stredná hodnota hodnôt priemernej ročnej teploty štandardného normálového obdobia je 9,556666667 °C a stredná hodnota hodnôt priemernej ročnej teploty obdobia rokov 1995 – 2008 je 10,38571429 °C. A keďže sa na základe výsledkov dvojvýberového *t*-testu s rovnosťou rozptylov tieto stredné hodnoty významne odlišujú, môžeme definitívne konštatovať, že obdobie rokov 1995 – 2008 bolo z hľadiska priemernej ročnej teploty vzduchu v porovnaní so štandardným normálovým obdobím teplejšie.

Globálne otepľovanie má vplyv na všeobecnú cirkuláciu atmosféry, čo sa v niektorých oblastiach Zeme môže prejaviť dynamickejšími prechodmi tlakových útvarov. To môže viesť, ako uvádza [1], k častému striedaniu veľmi vysokých a nízkych teplôt vzduchu a tiež k zvýrazňovaniu teplotných extrémov. Pomocou tzv. jednovýberového Wilcoxonovho testu teda zistíme, či existuje významný rozdiel medzi variačnými šírkami priemernej mesačnej teploty vzduchu pre štandardné normálové obdobie (tabuľka 9) a variačnými šírkami priemernej mesačnej teploty vzduchu pre obdobie rokov 1995 – 2008 (tabuľka 9). Budeme testovať nulovú hypotézu H_0, ktorá tvrdí, že medzi variačnými šírkami priemernej mesačnej teploty oboch období nie je rozdiel, proti alternatívnej hypotéze H_1, ktorá je negáciou stanovenej nulovej hypotézy. Variačné šírky v danom období sú vypočítané ako rozdiel maximálnej a minimálnej hodnoty priemernej mesačnej teploty vzduchu pre príslušný mesiac.

Tabuľka 9: Var. šírky priem. mes. teploty vzduchu (Nitra – Janíkovce)

obdobie rokov	*mesiac*											
	I.	II.	III.	IV.	V.	VI.	VII.	VIII.	IX.	X.	XI.	XII.
	variačná šírka											
1961 - 1990	9,9	10,9	8,7	5,7	4,9	5,1	5,0	4,8	5,5	7,5	8,4	7,3
1995 - 2008	7,9	7,7	5,7	6,4	3,9	4,8	5,2	5,0	6,3	6,0	6,3	9,0

Zdroj: SHMÚ

Realizácia jednovýberového Wilcoxonovho testu podľa [3]:

1,5. 1,5. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11.

$|-0{,}2| \leq |-0{,}2| < |0{,}3| < |-0{,}7| < |-0{,}8| < |1{,}0| < |1{,}5| < |-1{,}7| < |2{,}0| < |2{,}1| < |3{,}0| < |3{,}2|$

Nech R_i je poradie hodnoty x_i v danej postupnosti absolútnych hodnôt rozdielov variačných šírok. Definujme nasledujúce štatistiky:

$$R^+ = \sum_{i:x_i \geq 0} R_i \qquad R^- = \sum_{i:x_i < 0} R_i \tag{1}$$

Testovacím kritériom je štatistika:

$$T_w = \min\left\{R^+, R^-\right\} \tag{2}$$

Konkrétne v našom prípade platí: $R^+ = 2 + 5 + 6 + 8 + 9 + 10 + 11 = 51$ a $R^- = 1{,}5 + 1{,}5 + 3 + 4 + 7 = 17$. Potom $T_w = \min\left\{R^+, R^-\right\} = \min\{51;17\} = 17$. Následne zistíme, či hodnota testovacieho kritéria T_w leží v kritickom obore W_α (3) s hladinou významnosti α.

$$W_\alpha = \left[0; W_n(\alpha)\right] \tag{3}$$

Kritické hodnoty $W_n(\alpha)$ nájdeme v tabuľke kritických hodnôt jednovýberového Wilcoxonovho testu, pričom n je počet členov neklesajúcej postupnosti. V našom prípade $n = 12$. Potom - ak zvolíme hladinu významnosti $\alpha = 0{,}05$ - dostávame: $W_\alpha = [0; W_{12}(0{,}05)] = [0;13]$. Ak hodnota testovacieho kritéria T_w leží v kritickom obore W_α, potom hypotézu H_0 zamietame a prijímame hypotézu H_1. V našom prípade vidíme, že $T_w \notin W_\alpha$, resp. $17 \notin [0;13]$, preto hypotézu H_0 nezamietame. Medzi variačnými šírkami priemernej mesačnej teploty vzduchu štandardného normálového obdobia a variačnými šírkami priemernej mesačnej teploty vzduchu obdobia rokov 1995 – 2008 sa teda nepreukázal významný rozdiel, z čoho vyplýva, že úvahy o zvýrazňovaní teplotných extrémov z hľadiska priemernej mesačnej teploty vzduchu podmienenom globálnym otepľovaním sú pre oblasť Nitry neopodstatnené.

3. Výsledky štatistickej analýzy

Najdôležitejším výsledkom, ku ktorému sme dospeli, je, že obdobie rokov 1995 – 2008 bolo ako celok v oblasti Nitry vzhľadom na štandardné normálové obdobie štatisticky významne teplejšie (približne o 0,8 °C). Preukázali sme to prostredníctvom dvojvýberového *t*–testu s rovnosťou rozptylov, ktorý sme aplikovali na hodnoty priemernej ročnej teploty vzduchu štandardného normálového obdobia a obdobia rokov 1995 – 2008. Iste, ak porovnáme oteplenie o 0,8 °C s normálnymi výkyvmi teploty medzi dňom a nocou alebo medzi dvomi dňami, nejaví sa to ani zďaleka ako veľká zmena. Ak však uvážime, že ide o zvýšenie hodnôt priemernej ročnej teploty vzduchu a navyše po uplynutí krátkeho časového obdobia, je to zmena výrazná.

Po analýze početností výskytu teplotne nadpriemerných a teplotne podpriemerných mesiacov prislúchajúcich k jednotlivým rokom obdobia 1995 – 2008 sme vyslovili hypotézu, že oteplenie v tomto období bolo postupné. Táto hypotéza sa aplikáciou regresnej analýzy na početnosti výskytu teplotne nadpriemerných mesiacov s cieľom určiť trend ich rastu potvrdila a podľa [1] tento výsledok zodpovedá jednému zo znakov globálneho otepľovania (postupné otepľovanie).

Štatisticky významné rozdiely vo variačných šírkach priemernej mesačnej teploty vzduchu medzi obdobím rokov 1995 – 2008 a štandardným normálovým obdobím sa na základe jednovýberového Wilcoxonovho testu nepreukázali, teda výkyvy v priemerných mesačných teplotách v závislosti od jednotlivých rokov boli v oboch obdobiach podobné. Najväčšie variačné šírky mala priemerná mesačná teplota vzduchu v mesiacoch zimného polroka.

4. Záver

Na základe výsledkov, ku ktorým sme dospeli, možno konštatovať, že zmeny v teplote vzduchu v oblasti Nitry v období rokov 1995 – 2008 nastali a ak sa o globálnom otepľovaní hovorí ako o reálnom probléme, potom mu tieto zmeny zodpovedajú.

Meteorológovia a klimatológovia v hojnej miere využívajú matematickú štatistiku a jej metódy. Obdobne ani my pri sledovaní zmien v teplotných pomeroch v oblasti Nitry, sme sa nevyhli tomu, aby sme dospeli k určitým záverom využitím štatistických metód, čím sme poukázali na užitočnosť štatistiky ako matematickej vednej disciplíny aj v tejto oblasti.

5. Literatúra

[1] HOUGHTON, J. 1995. Globální oteplování. Oxford: Lion Publishing, 1995.

[2] TIRPÁKOVÁ, A. – MALÁ, D. 2007. Základy štatistiky pre pedagógov, psychológov a sociológov s popisom postupu práce v programe Excel. Nitra: FPV UKF, 2007, ISBN 978-80-8094-220-5.

[3] VRÁBELOVÁ, M. – MARKECHOVÁ, D. 2001. Pravdepodobnosť a štatistika. Nitra: FPV UKF, 2001.

Adresa autorov:

Lenčéš Peter, PaedDr., RNDr.
Tr. A. Hlinku 1
949 74 FPV UKF v Nitre
peter.lences@ukf.sk

Janka Melušová, PaedDr., PhD.
Tr. A. Hlinku 1
949 74 FPV UKF v Nitre
jmelusova@ukf.sk

Ku prognózovaniu účasti vo voľbách
Towards Predicting Attendance at Election

Ján Luha

Abstract: This article presents some possibilities for predicting attendance at election in Slovakia.

Key words: Election, predicting, attendance

Kľúčové slová: Voľby, prognózovanie, účasť

1. Úvod

V roku 2009 nás čaká niekoľko volieb. Predseda NR SR vyhlásil dňa 8. 1. 2009 termíny volieb:
prezidentských: 1. kolo: 21. 3. 2009 2. kolo: 4. 4. 2009,
doplňujúce do samospráv: 25. 4. 2009
do Euro parlamentu: 6. 6. 2009.

Termín volieb **do VÚC** zatiaľ stanovený nebol. Parlamentné a komunálne voľby majú termín rok 2010.

Očakávané výsledky zaujímajú nielen kandidátov a odborníkov - politológov, ale do značnej miery aj verejnosť. Odhady volebných výsledkov budú zrejme sledovaným "produktom" agentúr výskumu verejnej mienky.

V tejto súvislosti je dôležitá otázka, ako sa podarí odhadnúť účasť vo voľbách. Meranie účasti vo voľbách vo výskumoch verejnej mienky je dôležité pre lepšie odhady volebných výsledkov. V prípade signifikantnej odchýlky odhadu účasti vo voľbách nie je, po aplikácii príslušného filtra súbor respondentov, ktorí odpovedajú na meritórne otázky týkajúce sa volebných výsledkov, ***reprezentatívny,*** a tým sú ovplyvnené aj odhady výsledkov meritórnych otázok.

Je dobre známe, že počnúc rokom 1994 sa priamou otázkou o účasti vo voľbách nedarí volebnú účasť odhadovať a autor nepozná výsledky výskumov, ak takéto boli, ktoré by dokázali volebnú účasť dobre odhadovať.

Poznámka: Výsledky o účasti uvádzame v percentách - znak percenta v tabuľkách a grafoch neuvádzame.

2. Komparácia skutočnej účasti vo voľbách s ich odhadmi z výskumov verejnej mienky ÚVVM

V štúdii ÚVVM [5] "Slovensko pred parlamentnými voľbami 2006" je uvedený prehľad skutočnej účasti vo voľbách uskutočnených v SR od roku 1990. Nie sú tam ale zosumarizované výsledky výskumov verejnej mienky, ktoré zisťovali účasť v odpovedajúcich voľbách. Výsledky zo štúdie, ktoré uvádzame v tabuľkách, sú doplnené o údaje za účasť v parlamentných voľbách v roku 2006 a v komunálnych voľbách v roku 2006. Autor doplnil tabuľky o výsledky odhadov účasti z výskumov ÚVVM, ktoré boli realizované v termíne najbližšom ku konaniu príslušných volieb.

Odhady volebnej účasti z výskumov verejnej mienky, okrem už spomínaného javu, keď od roku 1994 je ***deklarovaná účasť*** respondentmi signifikantne nadhodnotená, prinášajú aj problém, ktorá hodnota je vlastne odhad volebnej účasti. Zvyčajne sa používajú dva typy škál

(s malými obmenami- vo výskumoch do EP v roku 2004 a do NR SR v roku 2006 nebola použitá priama otázka na účasť, ale pomocou variantov v škále odpovedí.).

- **A:** 1=určite sa zúčastním; 2=asi sa zúčastním; 3=asi sa nezúčastním; 4=určite sa nezúčastním; 5=ešte neviem,
- **B:** 1=určite sa zúčastním; 2=určite sa nezúčastním; 3=ešte neviem.

Ako odhad volebnej účasti v škále typu A môžeme uvažovať:
1) súčet percent 1+2; 2) súčet percent 1+2+5/2; 3) percento za 1

Ako odhad volebnej účasti v škále typu B môžeme uvažovať:
1) percento za 1; 2) súčet percent 1+2/2

Na komparáciu so skutočnými výsledkami vo voľbách využijeme obvykle používaný odhad, čo je prvý variant pri oboch typoch škál. Treba poznamenať, že parlamentné voľby v rokoch 1990 a 1992 a tiež komunálne voľby v roku 1990 sa mierne líšia od ďalších volieb - tie boli už v samostatnej SR.

Tabuľka 1: Účasť vo voľbách

Parlamentné voľby	jún 1990	jún 1992	sept 1994	sept 1998	sept 2002	jún 2006
Skutočná účasť	*95,4*	*84,2*	75,7	84,2	70,1	54,7
Deklarovaná účasť z výskumov ÚVVM	*83,1*	*83,1*	85,1	*/ÚVVM nezisťoval	77,6	67,8

Komunálne voľby	dec 1990	sept 1994	dec 1998	dec 2002	dec 2006
Skutočná účasť	*63,8*	52,4	54,0	49,5	47,7
Deklarovaná účasť z výskumov ÚVVM	*73,44*	79,1	*/ÚVVM nezisťoval	72,7	61,2

Voľby do VÚC	dec 2001, 1. kolo	dec 2001, 2. kolo	nov 2005, 1. kolo	dec 2005, 2.kolo
Skutočná účasť	26,0	22,6	18,0	11,1
Deklarovaná účasť z výskumov ÚVVM	64,9	*/ÚVVM nezisťoval	45,4	*/ÚVVM nezisťoval

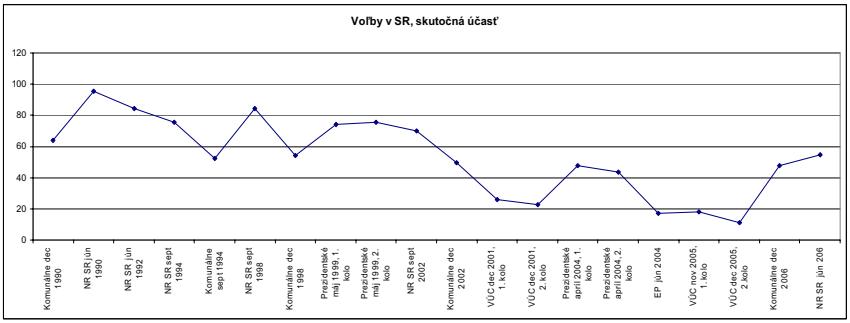
Prezidentské voľby	máj 1999, 1. kolo	máj 1999, 2. kolo	apríl 2004, 1. kolo	apríl 2004, 2. kolo
Skutočná účasť	73,9	75,5	47,9	43,5
Deklarovaná účasť z výskumov ÚVVM	87,1	87,1	74,2	*/ÚVVM nezisťoval

Voľby do EP	jún 2004
Skutočná účasť	17,0
Deklarovaná účasť z výskumov ÚVVM	57,6

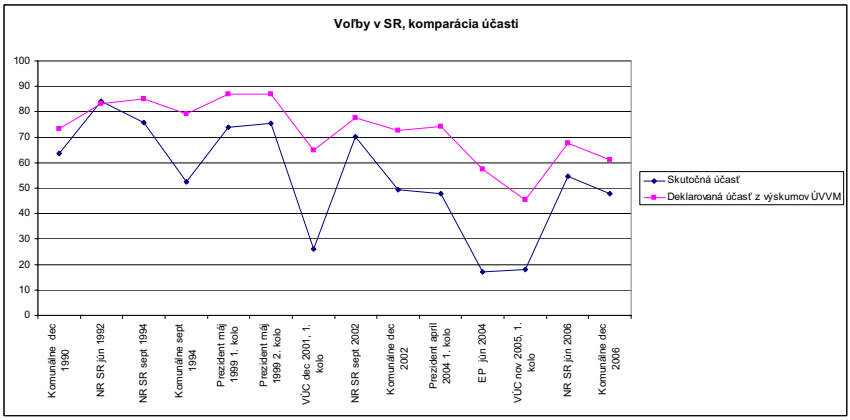
Zdroj [5]

V grafoch prehľadne uvádzame "priebeh" skutočnej účasti vo všetkých voľbách v SR od roku 1990, bez ohľadu na ich "typ", a komparáciu skutočnej účasti s deklarovanou v relevantných výskumoch verejnej mienky ÚVVM. zaujímavý je aj "priebeh" odchýlky odhadov volebnej účasti z výskumov verejnej mienky od skutočnej účasti vo voľbách. Treba poznamenať, že grafy sú ilustratívne nielen pretože zaznamenávajú rôzne typy volieb ale tiež aj rôzne intervaly medzi nimi.

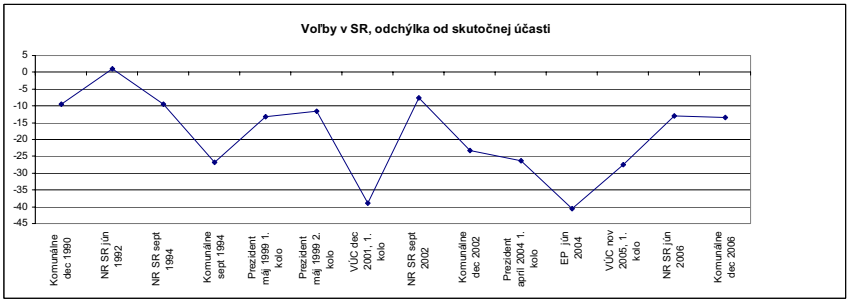
Poznámka: Od konca roka 1994 až do začiatku roka 1999 ÚVVM volebné preferencie nezisťoval.



Obrázok 1: Voľby v SR, skutočná účasť



Obrázok 2: Voľby v SR, skutočná účasť



Obrázok 3: Voľby v SR, odchýlka od skutočnej účasti

3. Aspekty prognóz volebnej účasti

Ako vidno z prehľadu skutočnej účasti vo voľbách - je účasť (do istej miery) podmienená typom volieb. Ďalší evidentný fakt je znižovanie účasti vo voľbách. Účasť v referendách je špecifická téma - v tomto príspevku sa ňou nezaoberáme.
V roku 1994 sme zaznamenali prvýkrát signifikantné nadhodnotenie odhadu účasti v parlamentných voľbách z výskumov verejnej mienky a odvtedy sa toto nadhodnotenie ešte zvýšilo. Preto sa jednoduchou otázkou na účasť vo voľbách táto veličina merať nedá.
Odhad účasti vo voľbách sa môžeme pokúsiť realizovať na základe výskumov verejnej mienky, na základe skutočnej účasti, pomocou expertných odhadov alebo iných kvalitatívnych výskumov, prípadne ich kombináciou.

Prognózy volebnej účasti vo výskumoch verejnej mienky:
Pri výskumoch verejnej mienky je dôležité pýtať sa reprezentatívnej množiny respondentov, teda by bolo vhodné vytvoriť vhodný nástroj na dobré odhadovanie účasti, resp. neúčasti vo voľbách. Takýto odhad by mal lepšie "separovať" voličov a nevoličov a tým by sa mohol zlepšiť aj odhad výsledkov otázky "koho bude voliť". Preto je potrebné ,okrem priamej otázky účasti vo voľbách, skonštruovať indikátory, ktoré takýto odhad pomôže skonštruovať. Jedným z takýchto indikátorov by mohla byť otázka o "pevnosti" rozhodnutia zúčastniť sa volieb. Autor príspevku nepozná úspešné pokusy vytvoriť vhodný model na spresnenie odhadov účasti, preto uvádzame iba niekoľko poznámok ako by sa mohlo postupovať:

- skupina odborníkov navrhne systém indikátorov, na základe ktorých sa experimentálne overí model prognózy účasti vo voľbách,
- indikátory budú vygenerované na základe kvalitatívnych výskumov a tiež overíme model prognózy účasti vo voľbách,
- kombinácia oboch postupov.

Prognózy volebnej účasti pomocou kvalitatívnych výskumov:
Kvalitatívne výskumy (napríklad expertné metódy) môžu slúžiť aj ako nástroj na odhad samotnej účasti vo voľbách.

Prognózy volebnej účasti na základe skutočnej účasti:
Skutočne zistené výsledky účasti vo voľbách môžu slúžiť na odhad účasti vo voľbách. Ilustratívne príklady uvedieme v 4. časti. Taktiež sa možno pokúsiť využiť prípadnú koreláciu medzi skutočnosťou a výsledkami z výskumov verejnej mienky.

4. Náčrt odhadov volebnej účasti

Prognostika poskytuje celú škálu metód, napríklad regresné metódy, odhady založené na poslednej skutočnosti - tie sú ale vhodnejšie pri krátkodobých predpovediach. Možno využiť aj kvalitatívne výskumy - z nich sú pravdepodobne najvhodnejšie expertné metódy ako napr. Delphi forecast method. Prognostické metódy vyžadujú relevantné dáta, čo je pri volebných prognózach v SR problém.

Ako prvý odhad účasti vo voľbách v roku 2009 použijeme naposledy zistenú skutočnú účasť - i keď je "vzdialenosť" medzi voľbami z hľadiska prognóz dosť značná.

Odhady sú v tabuľkách 2 a 3. Po ďalšej analýze, aj pomocou nástrojov navrhnutých v 3. časti, môžu byť odhady (prognózy) spresňované.

Tabuľka 2: Odhady účasti na voľbách v r. 2009

Prezidentské voľby 2009	**1. kolo**	**2. kolo**
Prvý odhad účasti	47,9	43,5
Voľby do EP 2009	**Prvý odhad účasti**	
	17,0	
Voľby doVÚC 2009	**Prvý odhad účasti**	
	11,1	

Tabuľka 3: Odhady účasti na voľbách v r. 2010

Komunálne voľby 2010	**Prvý odhad účasti**
	47,7
Parlamentné 2010	Prvý odhad účasti
	54,7

Ďalšie "náznaky" prognóz účasti vo voľbách v SR v roku 2009.

Ako prvé budú v roku 2009 voľby prezidenta Slovenskej republiky. ***Prezidentské voľby*** priamou voľbou prezidenta občanmi boli v histórii SR dvakrát a tak na prognózovanie napríklad pomocou regresných modelov nemáme dostatok údajov. Pokles skutočnej účasti bol: v 1. kole v roku 1999 zo 73,9% na 47,9% v 1. kole v roku 2004 a podobne v 2. kole bol pokles zo 75,5% na 43,5%. V období volieb pôsobia rozličné faktory vyplývajúce z politickej a ekonomickej situácie. Vplyv týchto faktorov je veľmi ťažké zapracovať do modelu prognózy účasti v prezidentských voľbách, preto priame "predĺženie" regresnej priamky ***nepovažujeme za dobrý odhad účasti***. I tak si tento odhad zapíšeme.
Druhý odhad účasti v prezidentských voľbách v roku 2009 je teda:

1. kolo	**2. kolo**
21,9	11,5

A teraz už môžeme radostne čakať na skutočné výsledky 21. marca a v prípade potreby druhého kola aj 4. apríla.

Prognóza účasti ***vo voľbách do EP*** je zrejme tvrdým orieškom aj pre expertov. Na základe jednej hodnoty 17,0% z roku 2004 a okolností, ktoré môžu pôsobiť v období pred voľbami začiatkom júna 2009, sa ťažko predpovedá.

Odhad účasti vo voľbách do VÚC je podobným orieškom ako pri prezidentských voľbách. Účasť v 1. kole 26,0% v roku 2001 poklesla na 18% v roku 2005 a v 2. kole poklesla z 22,6% na 11,1%. Keď aplikujeme analogický postup ako pri prezidentských voľbách dostávame druhý odhad účasti vo voľbách do VÚC v roku 2009:

1. kolo	**2. kolo**
10	-0,4

S takýmto odhadom ale asi ťažko možno súhlasiť.

Na odhady účasti vo voľbách v roku 2010 je ešte čas, ale iné údaje už nebudú k dispozícii, preto môžeme realizovať druhý pokus odhadu účasti v parlamentných a komunálnych voľbách.

Na prognózu parlamentných volieb máme 6 údajov o skutočnej účasti. V samostatnej SR boli voľby počnúc od roku 1994 a tak na prognózu použijeme 4 údaje z nich. Mierne zvýšenie účasti v parlamentných voľbách v roku 1998 môžeme pravdepodobne prisúdiť "protimečiarovskému" ťaženiu. Keďže interval medzi týmito 4 voľbami je približne rovnaký na druhý odhad účasti použijeme lineárnu regresnú krivku. Z výsledkov lineárnej regresie uvedenej v tabuľke 4 ***dostávame odhad účasti v roku 2010: 51,9%.*** Vzhľadom na malý počet údajov je ale regresný model nesignifikantný P=0,2.

Tabuľka 4: Odhady účasti na parlamentných voľbách v r. 2010 (regresný model)

Model Summary

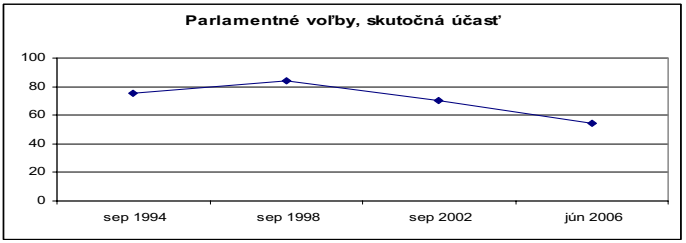
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,801(a)	,642	,464	9,09635

a Predictors: (Constant), porcis

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	90,450	11,141		8,119	,015
	porcis	-7,710	4,068	-,801	-1,895	,199

a Dependent Variable: ucast



Údaje o skutočnej účasti v 5 voľbách máme k dispozícii ***na prognózu komunálnych volieb***. Medzi voľbami bol približne rovnaký 4 ročný interval. Krivka účasti v komunálnych voľbách naznačuje možnosť použiť na prognózu lineárnu regresiu. I keď je charakter volieb v roku 1990 iný - v samostatnej SR boli 4 komunálne voľby, prognóza pomocou lineárnej regresie je takmer rovnaká, preto použijeme všetkých 5 údajov. Z výsledku lineárnej regresie dostávame ***druhý odhad účasti 43,0%.*** Model je signifikantný, keďže P=0,046.

Tabuľka 5: Odhady účasti na komunálnych voľbách v r. 2010 (regresný model)

Model Summary

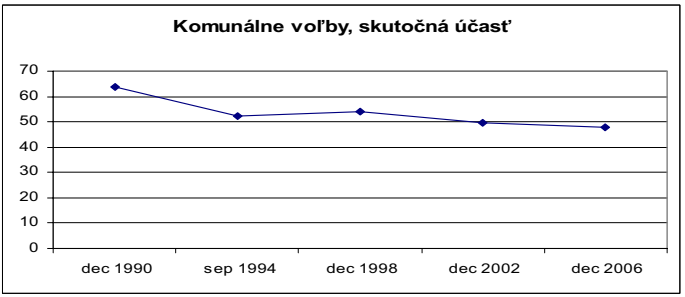
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,885(a)	,784	,712	3,3659

a Predictors: (Constant), volby

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	64,010	3,530		18,132	,000
	volby	-3,510	1,064	-,885	-3,298	,046

a Dependent Variable: ucast



5. Namiesto záverov

Namiesto záverov dve poznámky:

1. Prognózy volebnej účasti necháme na expertov.
2. Odhady volebnej účasti konštruované na základe skutočnej účasti, poprípade na základe expertných odhadov, nepomôžu pri odhadovaní účasti z výskumov verejnej mienky.

6. Literatúra

[1] http://portal.statistics.sk/showdoc.do?docid=30
[2] CHAJDIAK, J.: Štatistické úlohy a ich riešenie v Exceli. STATIS, Bratislava 2005, ISBN 80- 85659-39-5.
[3] KANDEROVÁ, M. – ÚRADNÍČEK, V.: Štatistika a pravdepodobnosť pre ekonómov. 1. časť. OZ Financ, Banská Bystrica 2007, ISBN 978-80-969535-5-4.
[4] KANDEROVÁ, M. – ÚRADNÍČEK, V.: Štatistika a pravdepodobnosť pre ekonómov. 2. časť. OZ Financ, Banská Bystrica 2007, ISBN 987-80-696535-6-1.
[5] Kol.: Slovensko pred parlamentnými voľbami 2006. Štatistický úrad SR, Verejná mienka, jún 2009
[6] LUHA J.: Reprezentatívnosť vo výskumoch verejnej mienky. FORUM STATISTICUM SLOVACUM 2/2005. ISSN 1336-7420. SŠDS Bratislava 2005.
[7] LUHA J.: Kvótový výber. FORUM STATISTICUM SLOVACUM 1/2007. SŠDS Bratislava 2007. ISSN 1336-7420.
[8] LUHA J.: Korelácie disjunktných a komplementárnych javov. FORUM STATISTICUM SLOVACUM 6/2008. SŠDS Bratislava 2008. ISSN 1336-7420.
[9] PECÁKOVÁ I.: Statistika v terénních průzkumech. Proffessional Publishing, Praha 2008. ISBN 978-80-86946-74-0.
[10] ŘEZANKOVÁ A.: Analýza dat z dotazníkových šetření. Proffessional Publishing, Praha 2007. ISBN 978-80-86946-49-8.
[11] STANKOVIČOVÁ I., VOJTKOVÁ M.: Viacrozmerné štatistické metódy s aplikáciami. IURA EDITION, BRATISLAVA 2007, ISBN 978-80-8078-152-1.
[12] Výsledky relevantných výskumov ÚVVM publikované pred jednotlivými voľbami v "Informáciách pre tlač".

Adresa autora:

Luha Ján, RNDr., CSc.
Ústav pre výskum verejnej mienky pri ŠÚ SR
Hanulova 5/C
841 01 Bratislava
Jan.Luha@statistics.sk

O Benfordovom zákone
About Benford´s Law

Dagmar Markechová, Jaroslav Ivor, Anna Tirpáková, Beáta Stehlíková

Abstract: The contribution describes Benford´s law and its application in practice. We introduce the programme in software Mathematica which computes the expected frequencies of the digits by Benford's law. These theoretical frequencies are compared with empirical frequencies of the digits.

Key words: Benford´s law, programme Mathematica, empirical frequencies of the digits, expected frequencies of the digits, χ^2 – test

Kľúčové slová: Benfordov zákon, program Mathematica, empirické početnosti, očakávané početnosti, χ^2 – test

1. Úvod

Benfordov zákon je matematický zákon, ktorý hovorí, že vo viacerých súboroch prirodzených čísel (ale nie vo všetkých) čísla oveľa častejšie začínajú číslicou 1 ako inými číslicami. Približne 30 % čísel začína jedničkou. Čím je vyššia číslica, tým je menšia pravdepodobnosť, že sa na začiatku čísla objaví. Benfordov zákon platí pre údaje z reálneho sveta, ale aj pre sociálne významné údaje. Súbory čísel môžu predstavovať napríklad čiastky na účtoch za elektrinu, počty novinových článkov, popisné čísla, ceny na burzách, veľkosti populácií, veľkosti povodí a dĺžky riek alebo fyzikálne a matematické konštanty. Pretože ľudia majú tendenciu pri vymýšľaní falošných výsledkov používať na začiatku čísel všetky číslice s rovnakou pravdepodobnosťou, Benfordov zákon sa môže využívať napríklad v kriminalistike pri hľadaní možných podvodov [10] alebo pri jednoduchom testovaní, či voľby boli regulárne [11]. Pre signifikantné overovanie platnosti zákona stačí mať výsledky v rozsahu aspoň troch rádov.

Ako prvý o tomto zákone publikoval v roku 1881 americký astronóm Simon Newcomb, ktorý si všimol, že logaritmické tabuľky sú viac používané na stránkach, kde sa vyskytujú čísla začínajúce jedničkou. Zákon bol znovu objavený v roku 1938 fyzikom Frankom Benfordom, podľa ktorého sa dnes nazýva.

Zjednodušene Benfordov zákon hovorí, že začiatočná číslica n $(n \in \{1,...,b-1\})$ čísla v číselnej sústave so základom b $(b \geq 2)$ sa objavuje s pravdepodobnosťou

$$\log_b(n+1) - \log_b(n).$$

V desiatkovej sústave (b=10) majú číslice na prvom mieste čísla podľa Benfordovho zákona nasledujúce pravdepodobnostné rozdelenie[6]:

Tabuľka 1: Pravdepodobnostné rozdelenie výskytu číslic desiatkovej sústavy na prvom mieste čísla podľa Benfordovho zákona

1	2	3	4	5	6	7	8	9
30,10%	17,60%	12,50%	9,70%	7,90%	6,70%	5,80%	5,10%	4,60%

Pravdepodobnostné rozdelenie číslic na prvom mieste daného čísla je graficky znázornené na obrázku 1.

2. Materiál a metódy

Výpočet teoretického podielu jednotlivých cifier na 1., 2. až *i*-tom mieste čísel daného súboru je možné realizovať pomocou počítača, napríklad v programe Mathematica. Autori navrhujú nasledujúci program pre výpočet podielu jednotlivých cifier na 1. až ľubovoľnom i-tom (i = 2, 3, ...) mieste čísel daného súboru.

```
B = 10; qq = Table[Log[1. + 1/d]/Log[B], {d, 1, 9}];
qq2 = Table[{i, qq[[i]]}, {i, 1, 9}];
s = Sum[qq[[i]], {i, 1, 9}];
Print["Očakávaný podiel výskytu jednotlivých cifier 1 až 9 na prvom mieste podľa Bonferdovho zákona: "];
qq2 // TableForm

<< Graphics`Graphics`
BarChart[Table[{qq[[i]], i}, {i, 1, Length[qq2]}]];

k = 2; B = 10; n = k - 1; qq = Table[Sum[Log[1. + 1/(k*B + d)], {k, B^(n - 1), (B^n) - 1}]/Log[B], {d, 0, 9}];
qq2 = Table[{i - 1, qq[[i]]}, {i, 1, 10}];
s = Sum[qq[[i]], {i, 1, 10}];
Print["Očakávaný podiel výskytu jednotlivých cifier 1 až 9 na ", k, ". mieste podľa Bonferdovho zákona: "];
qq2 // TableForm

<< Graphics`Graphics`
BarChart[Table[{qq[[i]], i - 1}, {i, 1, Length[qq2]}]];
```

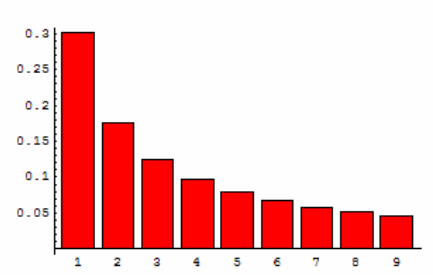
Použitím uvedeného programu pre i=5 sme dostali nasledovné výsledky, ktoré predstavujú očakávaný podiel výskytu jednotlivých cifier 1 až 9 na prvom a cifier 0 až 9 na druhom až piatom mieste podľa Benfordovho zákona.

Tabuľka 2: Očakávaný podiel výskytu jednotlivých cifier na prvom až piatom mieste podľa Benfordovho zákona

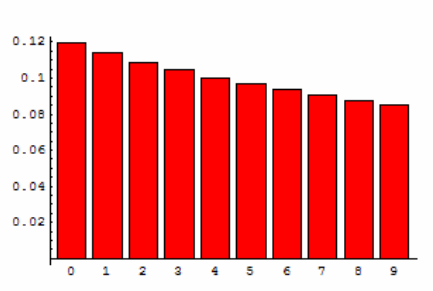
0		0.119679	0.101784	0.100176	0.100018
1	0.30103	0.11389	0.101376	0.100137	0.100014
2	0.176091	0.108821	0.100972	0.100098	0.10001
3	0.124939	0.10433	0.100573	0.100059	0.100006
4	0.09691	0.100308	0.100178	0.100019	0.100002
5	0.0791812	0.0966772	0.0997876	0.0999803	0.099998
6	0.0669468	0.0933747	0.0994013	0.0999412	0.0999941
7	0.0579919	0.090352	0.0990192	0.0999022	0.0999902
8	0.0511525	0.0875701	0.0986412	0.0998633	0.0999863
9	0.0457575	0.0849974	0.0982672	0.0998244	0.0999824

Výsledky, uvedené v predchádzajúcej tabuľke pre i = 1, 2, 3, sú graficky znázornené na nasledujúcich obrázkoch (Obrázok 1 až 3). Z obrázku, ktorý vyjadruje očakávaný podiel výskytu jednotlivých cifier 0 až 9 na treťom mieste čísel, je zrejmé, že ide o takmer

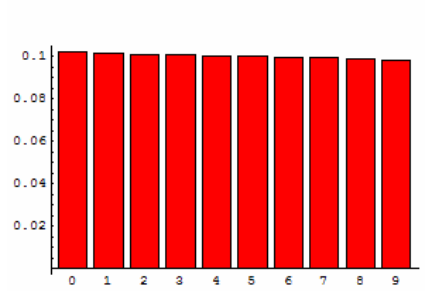
rovnomerné rozdelenie početností. Analogická situácia je pre funkciu rozdelenia početností na ďalších miestach v čísle (i = 4, 5, ...).



Obrázok 1: Očakávaný podiel výskytu jednotlivých cifier 1 až 9 na prvom mieste podľa Benfordovho zákona



Obrázok 2: Očakávaný podiel výskytu jednotlivých cifier 0 až 9 na druhom mieste podľa Benfordovho zákona



Obrázok 3: Očakávaný podiel výskytu jednotlivých cifier 0 až 9 na treťom mieste podľa Benfordovho zákona

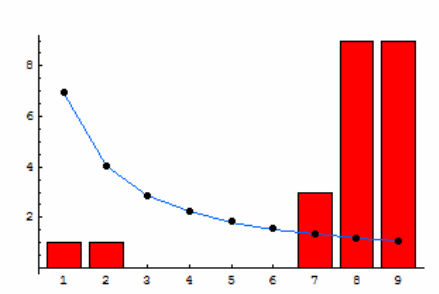
3. Výsledky a diskusia

V praxi nestačí posudzovať iba na základe obrázku, či údaje spĺňajú Benfordov zákon, ale je potrebné túto hypotézu štatisticky testovať. V nasledujúcom nami navrhnutom programe v programe Mathematica je použitý χ^2 – test dobrej zhody [1] empirického a teoretického Benfordovho pravdepodobnostného rozdelenia jednotlivých cifier na zvolenom mieste. Navrhnutý program sme aplikovali pre súbor čísel, ktorý sme dostali pomocou generátora náhodných čísel.

```
Print["TESTOVANIE ZHODY ROZDELENIA 1. CIFIER S BENFORDOVÝM ZÁKONOM "]
<< Statistics`ContinuousDistributions`
(* 1. cifra *)
rudaje = {};
udaje = {19, 27, 86, 72, 81, 97, 93, 89, 87, 92, 79, 87, 96, 91, 84, 90, 93,
   88, 94, 83, 96, 88, 71};

For[i = 1, i ≤ Length[udaje], i++,
  rudaje = AppendTo[rudaje, StringTake[ToString[udaje[[i]]], 1]];
 ];
q = Table[Count[rudaje, ToString[i]], {i, 1, 9}]; q2 = Table[{i, q[[i]]}, {i, 1, 9}];
Print["Skutočné početnosti výskytu jednotlivých cifier 1 až 9 na prvom mieste: "];
                                                          q2 // TableForm

B = 10; qq = Table[Log[1. + 1/d] * Length[udaje] / Log[B], {d, 1, 9}];
qq2 = Table[{i, qq[[i]]}, {i, 1, 9}];
Print["Očakávané početnosti výskytu jednotlivých cifier 1 až 9 na prvom mieste: "];
w = (q - qq)^2 / qq ; chi2 = Sum[w[[i]], {i, 1, 9}];
KH = Quantile[ChiSquareDistribution[8], 0.95] ; Phodnota = 1 - CDF[ChiSquareDistribution[8], chi2];
                                                                qq2 // TableForm

Print["Hodnota testovacej štatistiky je ", chi2, "; prílušná P-hodnota je ", Phodnota]
If[chi2 > KH, "ZÁVER: Prvé cifry údajov nie sú v súlade s Benfordovým zákonom", "ZÁVER:
                                  Prvé cifry údajov sú v súlade s Benfordovým zákonom"]

<< Graphics`Graphics`
a = BarChart[q];
b = ListPlot[qq2, PlotJoined → True, PlotStyle → Hue[.6]];
c = ListPlot[qq2, PlotStyle → PointSize[0.02], PlotStyle → Hue[.6]];
Show[a, b, c]
```

Výsledok testu pre podiel výskytu jednotlivých cifier 1 až 9 na prvom mieste čísel je nasledovný.

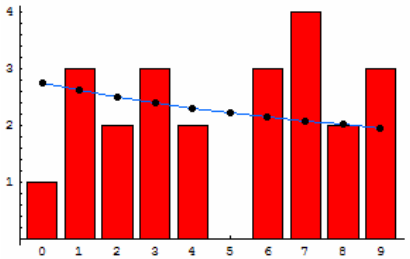
Hodnota testovacej štatistiky je 129,952; príslušná P-hodnota je 0,0000. To znamená, že testovanú hypotézu zamietame v prospech alternatívnej hypotézy. Test potvrdil, že pravdepodobnostné rozdelenie číslic na prvom mieste údajov nie je v súlade s Benfordovým zákonom. Získaný výsledok je graficky znázornený na obrázku 4.



Obrázok 4: Očakávaný a empirický podiel výskytu jednotlivých cifier 1 až 9 na prvom mieste podľa Benfordovho zákona

Výsledok testu pre podiel výskytu jednotlivých cifier 0 až 9 na druhom mieste čísel je nasledovný:

Hodnota testovacej štatistiky je 6,36145; príslušná P-hodnota je 0,703269. To znamená, že testovanú hypotézu nemôžeme zamietnuť v prospech alternatívnej hypotézy. Test potvrdil, že pravdepodobnostné rozdelenie číslic na druhom mieste údajov je v súlade s Benfordovým zákonom. Pozorované rozdiely medzi teoretickým a empirickým rozdelením nie sú štatisticky významné. Získaný výsledok je graficky znázornený na obrázku 5.



Obrázok 5: Očakávaný a empirický podiel výskytu jednotlivých cifier 0 až 9 na druhom mieste podľa Benfordovho zákona

4. Záver

V predloženom článku je prezentovaný Benfordov zákon a jeho praktické použitie je ilustrované na príklade. Pri riešení boli použité nami navrhnuté programy v softvéri Mathematica. Tieto programy sú univerzálne, možno ich použiť v praxi, napríklad na odhalenie finančných podvodov, legalizácie príjmov z trestnej činnosti a podobne.

5. Literatúra

[1]ANDEL, J.: Matematická statistika. Praha: SNTL, Alfa, 1985

[2]BROWN, M. W.: "Following Benford's Law, or Looking Out for No. 1", (From The New York Times, Tuesday, August 4, 1998

[3]HILL, T. P.: The First-Digit Phenomenon. , American Scientist, July-August 1998

[4]NIGRINY, M. J. – MITTERMAIER, L.I.: The Use of Benford's Law as an Aid in Analytical Procedures, Auditing: A Journal of Practice and Theory 16 (Fall): 5267, 1997.

[5]NIGRINY, M. J.: Using Digital Frequencies to Detect Fraud. The White Paper (April/May): 36, 1996

[6]RAIMI, R.: The First Digit Problem. American Mathematical Monthly 83 (Aug.Sept.): 521538, 1976

[7] HILL, T.P.: The First-Digit Phenomenon. , American Scientist, July-August 1998

[8] http://cs.wikipedia.org/wiki/Benford%C5%AFv_z%C3%A1kon

[9] http://www.ezrstats.com/Benford2.htm

[10] http://www.intuitor.com/statistics/Benford's%20Law.html

[11] http://www.aicpa.org/pubs/jofa/aug2003/rose.htm

[12] http://www.exceluser.com/tools/benford_xl11.htm

[13] http://mathworld.wolfram.com/BenfordsLaw.html

Adresy autorov:

Dagmar Markechová, Doc. RNDr., CSc.
Katedra Matematiky UKF
Tr. A. Hlinku 1
949 01 Nitra1
atirpakova@ukf.sk

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

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

Beáta Stehlíková, Prof. RNDr., CSc.
Fakulta ekonómie a podnikania BVŠP
Tematínska 10
851 05 Bratislava
biostateduca2003@yahoo.com

Objavovanie znalostí na základe používania webu – metódy a aplikácie
Web Usage Mining-Methods and Applications

Michal Munk

Abstract: In the article we deal with web usage mining and methods regarding to Internet specifics. In the paper we derive sequence rules from common association rules. In the conclusion, we indicate certain possibilities to apply found rules in prediction of visits sequence of different web parts and describe own experience with methods application and suggest practices for problem solving from web usage mining.

Key words: Web usage mining, association rules, sequence rules, web optimalization, adaptive hypermedia systems.

Kľúčové slová: Objavovanie znalostí na základe používania webu, asociačné pravidlá, sekvenčné pravidlá, optimalizácia webu, adaptívne hypermediálne systémy.

1. Objavovanie znalostí

V dnešnej dobe sa zautomatizoval zber dát a vznikla potreba tieto dáta používať v procese rozhodovania. Všetky tieto údaje sú dôležité na tvorbu rôznych akcií, na predvídanie správania zákazníkov, zisťovanie príčin porúch a pod. Obrovské množstvo údajov má ale slabú vypovedajúcu hodnotu. Za týmto účelom vznikla oblasť objavovania znalostí z databáz (Knowledge Discovery in Databases, KDD). KDD chápeme ako proces, ktorý zahrňuje výber dát, predspracovanie dát, transformáciu dát, analýzu dát a interpretáciu výsledkov. Na druhej strane sú tu aj iné zdroje dát ako samotné databázy a dátové sklady. Príkladom významného zdroja dát je samotný Internet, ktorý v súčasnosti predstavuje najdynamickejšie sa rozvíjajúci zdroj informácii. Z potreby analyzovať tieto dáta vznikla príbuzná oblasť KDD – objavovanie znalostí z webu (web mining), ktorá rieši svoje špecifické problémy. Niekedy stačí mierne prispôsobiť existujúce postupy z oblasti KDD, niekedy je potrebné zásadnejšie zmeniť kroky predspracovania a transformácie dát. Vo väčšine prípadov je možné použiť štandardné metódy data miningu a inokedy sa aplikujú celkom nové metódy zohľadňujúce zvláštnosti Internetu.

Rozlišujeme nasledovné úlohy objavovania znalostí z webu [1]:

- objavovanie znalostí na základe obsahu webu (web content mining)
- objavovanie znalostí na základe štruktúry webu (web structure mining)
- objavovanie znalostí na základe používania webu (web usage mining)

Cieľom objavovania znalostí na základe používania webu je analýza správania používateľa pri prechádzaní webu [2], [3]. Napríklad, zdrojom dát môžu byť automaticky ukladané dáta v logovacích súboroch (zachytávajú informácie o IP adrese používateľovho počítača, dátume a čase prístupu, navštívenom URL a pod.). V takýchto dátach sledujeme postupnosti – sekvencie v návšteve jednotlivých stránok používateľom, ktorého identifikuje IP adresa. V sekvenciách môžeme hľadať vzorce správania sa používateľov. Za týmto účelom je najlepšie použiť sekvenčnú analýzu, ktorej cieľom je extrakcia sekvenčných pravidiel. Prostredníctvom týchto pravidiel predikujeme sekvencie návštev rôznych webových častí používateľom. Táto metóda bola odvodená z asociačných pravidiel a je príkladom metódy zohľadňujúcej zvláštnosti Internetu.

Príkladom použitia štandardných metód data miningu pri analýze prístupu k webu môžu byť asociačné pravidlá a zhluková analýza. Napríklad, ak obchodné reťazce vydávajú nákupné karty, za účelom získania informácií o zákazníkoch (prostredníctvom ktorých si

môžu uplatňovať zľavy), čím získajú evidenciu zákazníkov spolu s väzbou na ich nákupy, potom o to jednoduchšie to musí byť pre elektronické obchody, ktoré získajú informácie pri registrácii zákazníka. Segmentáciou, napríklad použitím zhlukovej analýzy, môžeme skúmať správanie skupín klientov a následne získané segmenty popísať asociačnými pravidlami.

Cieľom príspevku je prezentovať sekvenčnú analýzu ako príklad metódy zohľadňujúcej zvláštnosti Internetu. Sekvenčné pravidlá odvodzujeme zo známejších asociačných pravidiel. V závere naznačujeme možnosti aplikácie nájdených pravidiel pri predikovaní sekvencie návštev rôznych webových častí používateľom a popisujeme vlastné skúsenosti s aplikáciou metód a navrhujeme postupy na riešenie problémov z oblasti objavovania znalostí na základe používania webu.

2. Asociačné pravidlá

Asociačné pravidlá sú jednou z najpopulárnejších metód hĺbkovej analýzy (data miningu). Úspešne sa aplikujú v analýze nákupného košíka, finančných dát, cenzusu a pod. Asociačné pravidlá nám umožňujú získať obraz o vzťahoch medzi prvkami v danej množine dát, pričom výsledky sa veľmi ľahko interpretujú. Pravidlá totiž predstavujú konštrukciu IF THEN, ktorú nájdeme vo všetkých programovacích jazykoch a dajú sa vyjadriť v prirodzenom jazyku. Asociačné pravidlá spopularizoval začiatkom 90. rokov Agraval v súvislosti s analýzou nákupného košíka [4]. Asociačné pravidlá sa radia medzi symbolické metódy strojového učenia a rovnako ako stromy sú šité na mieru kvalitatívnym/kategoriálnym dátam. Kvantitatívne/numerické dáta vyžadujú diskretizáciu - nahradenie numerických hodnôt intervalmi hodnôt, s ktorými sa potom narába ako s kategorickou premennou.

Nasledujúca časť ponúka teoretický rozbor asociačných pravidiel z matematického hľadiska.

Nech $D = \{T_1, T_2, \ldots, T_n\}$ je množina n transakcií a nech I je množina položiek, $I = \{i_1, i_2, \ldots, i_m\}$. Každá transakcia je množinou položiek, t. j. $T_i \subseteq I$. Asociačné pravidlo je implikácia v tvare $X \Rightarrow Y$, kde $X, Y \subset I$, a $X \cap Y = \emptyset$; X sa nazýva predpoklad/podmienka (antecedent/body) a Y záver/následok (consequent/head) pravidla. Vo všeobecnosti, množina položiek, ako napríklad X alebo Y, sa nazýva položková množina (itemset).

Príkladom T_i môže byť jeden nákup, pričom D predstavuje celú databázu, t. j. všetky nákupy za určité časové obdobie. Výsledkom analýzy sú pravidlá tvaru AK podmienka POTOM následok (IF body THEN head). Pravidlá sú určované na základe početností, s akými sa podmienka a následok vyskytujú v dátach.

Príklad asociačného pravidla z analýzy nákupného košíka:

{chlieb, syr} $\Rightarrow$ {maslo}, *confidence* = 57%, *support* = 21%.

Interpretácia pravidla je nasledovná: Zákazníci, ktorí si na jeden nákup kúpia chlieb a syr si s 57% pravdepodobnosťou kúpia aj maslo, pričom 21% nákupov obsahovalo chlieb, syr aj maslo.

Nech $P(X)$ je pravdepodobnosť výskytu položkovej množiny X v D a nech $P(Y|X)$ je podmienená pravdepodobnosť výskytu položkovej množiny Y, za podmienky výskytu položkovej množiny X.

Pre položkovú množinu $X \subseteq I$, *support*(X) je definovaná ako časť transakcií $T_i \in D$ takých, že $X \subseteq T_i$, t. j. $P(X) = support(X)$. Podpora (*support*) pravidla $X \Rightarrow Y$ je definovaná ako $support(X \Rightarrow Y) = P(X \cup Y)$. Inak povedané pravidlo $X \Rightarrow Y$ má podporu s, ak s% transakcií v databáze obsahuje $X \cup Y$. V podstate podpora reprezentuje frekvenciu výskytu danej množiny položiek v databáze.

Ďalšou dôležitou charakteristikou asociačného pravidla $X \Rightarrow Y$ je miera reliability nazývaná spoľahlivosť (*confidence*), *confidence*($X \Rightarrow Y$) je definovaná ako $P(Y|X) = P(X \cup Y)/P(X) = support(X \cup Y)/support(X)$. Inak povedané pravidlo $X \Rightarrow Y$ má spoľahlivosť c, ak

c% transakcií v databáze, ktoré obsahujú položku *X*, taktiež obsahujú položku *Y*. Spoľahlivosť pravidla je pravdepodobnosť výskytu pravej strany pravidla za podmienky výskytu ľavej strany, t. j. je to percentuálny podiel pravidiel, ktorých ľavá strana je *X* a pravá *Y* zo všetkých, ktorých ľavá strana je *X*.

Ďalšou dôležitou charakteristikou asociačného pravidla $X \Rightarrow Y$ je miera zaujímavosti nazývaná zdvih (*lift*), *lift*($X \Rightarrow Y$) je definovaný ako *lift*($X \Rightarrow Y$) = $P(Y|X)/P(Y)$ = *confidence*($X \Rightarrow Y$)/*support*(Y). Táto miera určuje koľko krát častejšie sa *X* a *Y* vyskytujú spolu, než by to bolo, keby boli štatisticky nezávislé. Ak je miera *lift*($X \Rightarrow Y$) > 1 indikuje to, že sa *X* a *Y* vyskytujú častejšie spolu ako zvlášť.

Cieľom asociačnej analýzy je nájsť všetky pravidlá, ktorých miery sú väčšie alebo rovné ako špecifikovaná podpora (*minimum support*) a spoľahlivosť (*minimum confidence*) [4], [5]. Tieto dve miery vypovedajú o početnosti výskytu pravidla v databáze (*support*) a o sile pravidla (*confidence*). *K*- položková množina s podporou väčšou ako určené minimum sa nazýva frekventovaná/veľká/často opakujúca sa množina položiek. Celý proces získavania pravidiel, definovaný nižšie, pozostáva z dvoch krokov, kde sa najskôr získajú všetky frekventované množiny položiek, z ktorých sa potom vygenerujú samotné pravidlá.

Definícia 1

1. Daná je množina položiek *I*, vstup pozostáva z množiny transakcií *D*, kde každá transakcia *T* je neprázdna podmnožina položiek množiny *I*, $T \subseteq I$.

2. Daná je množina položiek $T \subseteq I$ a množina transakcií *D*, definujeme podporu *T* ako *support*(T) = $P(T)$.

3. Nastavením minimálnej podpory *min s*, kde $0 \leq min\ s \leq 1$, definujeme frekventované množiny položiek ako množiny položiek kde *support*(T) $\geq$ *min s*.

Definícia 2

1. Asociačné pravidlo je pár disjunktných množín položiek, predpokladu $X \subseteq I$ a záveru $Y \subseteq I$, kde $X \Rightarrow Y$ a $X \cap Y = \varnothing$.

2. Definujeme podporu asociačného pravidla $X \Rightarrow Y$ ako *support*($X \Rightarrow Y$) = $P(X \cup Y)$.

3. Definujeme spoľahlivosť asociačného pravidla $X \Rightarrow Y$ ako *confidence*($X \Rightarrow Y$) = $P(Y|X)$.

4. Nastavením minimálnej spoľahlivosti *min c*, kde $0 \leq min\ c \leq 1$, definujeme silné pravidlá ako pravidlá kde *confidence*($X \Rightarrow Y$) $\geq$ *min c*.

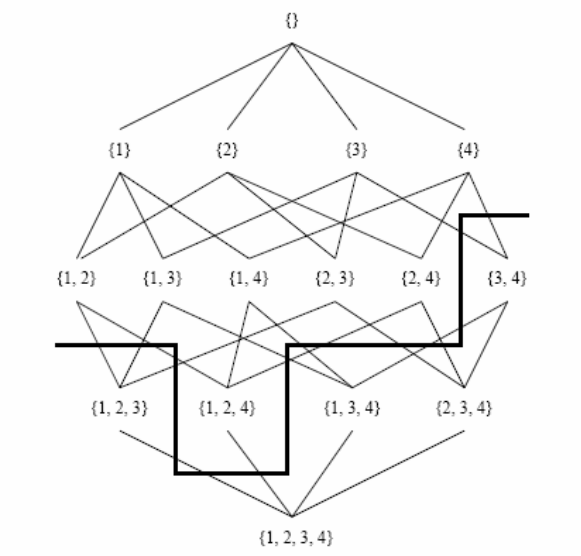
Najznámejší algoritmus Apriori navrhol Agrawal v súvislosti s analýzou nákupného košíka [6]. Algoritmus je založený na hľadaní už spomínaných frekventovaných položkových množín, ktoré predstavujú kombinácie/konjunkcie kategórií atribútov spĺňajúce podmienku minimálnej podpory.

Apriori algoritmus využíva generovanie kombinácii do šírky, t. j. najskôr sa vygenerujú kombinácie dĺžky jedna, potom všetky kombinácie dĺžky dva atd., pričom sa nesmú v kombinácii opakovať atribúty a položky sú usporiadané lexikograficky. Pri generovaní vlastne prehľadávame priestor všetkých prístupných kombinácií. Pri generovaní sa využíva vlastnosť frekventovaných položkových množín – každá podmnožina frekventovanej množiny položiek musí byť tiež frekventovaná, t. j. vo všeobecnosti na hľadanie kombinácii dĺžky *k*, sa využíva toho, že už poznáme (*k* - 1) - prvkové frekventované množiny.

Algoritmus nevyužívajúci túto vlastnosť frekventovaných množín by musel preskúmať všetkých 2^m možných podmnožín množiny položiek *I*, $I = \{i_1, i_2, \ldots, i_m\}$.

Uvedieme príklad [7], kde na obrázku 1 vidíme, že z kandidátov frekventovaných jednoprvkových množín *C*1 stanovenú minimálnu podporu spĺňajú všetky jednoprvkové množiny, t. j. *L*1 = {{1}, {2}, {3}, {4}}. Z tejto množiny je možné spájaním vygenerovať nasledovných kandidátov veľkosti 2, *C*2 = {{1, 2}, {1, 3}, {1, 4}, {2, 3}, {2, 4}, {3, 4}}. Minimálnu stanovenú hodnotu podpory spĺňajú nasledovné dvojprvkové množiny *L*2 = {{1,

2}, {1, 3}, {1, 4}, {2, 3}, {2, 4}}. Z uvedenej množiny dvojprvkových frekventovaných množín položiek možno spájaním vygenerovať nasledovných trojprvkových kandidátov $C3$ = {{1, 2, 3}, {1, 2, 4}, {1, 3, 4}, {2, 3, 4}}. Z tejto množiny orezávaním odstránime množiny {1, 3, 4}, {2, 3, 4}, vzhľadom na to, že ich podmnožina {3, 4} sa nenachádza v $L2$, t. j. nie je frekventovaná. Minimálnu stanovenú hodnotu podpory spĺňa iba nasledovná trojprvková množina $L3$ = {{1, 2, 4}}. Z tejto množiny trojprvkových frekventovaných množín, už ale nie je možné vygenerovať žiadneho štvorprvkového kandidáta, a tak algoritmus končí nájdením všetkých frekventovaných množín {$L1$, $L2$, $L3$} v D.



Zdroj: [7]

Obrázok 1: Hľadanie frekventovaných množín

Vo všeobecnosti by sme postup pri hľadaní frekventovaných množín mohli rozdeliť do troch krokov:

1. Vygenerovanie množiny kandidátov *Ck* na základe *Lk*-1 – spájanie.

2. Odstránenie takých množín z *Ck,* ktorých podmnožina sa nenachádza v *Lk*-1 – orezávanie.

3. Zaradenie takých množín z *Ck* do *Lk*, ktoré spĺňajú stanovenú minimálnu hodnotu podpory.

Po nájdení frekventovaných položkových množín, t .j. kombinácií, ktoré vyhovujú svojou početnosťou, sa vytvárajú pravidlá spĺňajúce podmienku minimálnej spoľahlivosti.

Každá kombinácia T sa rozdelí na všetky možné dvojice podkombinácií X a Y také, že $Y = T - X$. Na základe tejto skutočnosti podpora pravidla bude rovnaká ako podpora kombinácie, t. j. $support(X \Rightarrow Y) = support(T)$.

Spoľahlivosť vypočítame ako podiel početnosti kombinácie T a predpokladu X, $confidence(X \Rightarrow Y) = support(T)/support(X)$, pričom $support(X) \geq support(T)$, vzhľadom na to, že predpoklad X je skôr vygenerovaná frekventovaná položková množina, t. j. X je kratšia menej obmedzujúca kombinácia.

Ak *X'* je nadkombináciou kombinácie X, je *X'* prísnejšie, t. j. splní ho menej príkladov. Vzhľadom na to, že pri výpočte spoľahlivosti je početnosť *X'* v menovateli zlomku, pričom

čitateľ zostáva rovnaký, je *confidence*($X' \Rightarrow T - X'$) ≥ *confidence*($X \Rightarrow T - X$), t. j. ak nevyhovuje zadanej minimálnej spoľahlivosti pravidlo $X' \Rightarrow T - X'$, nevyhovuje ani žiadne pravidlo $X \Rightarrow T - X$.

Ak napr. pre kombináciu {1, 2, 4} pravidlo {1, 2}⇒ {4} nespĺňa podmienku minimálnej spoľahlivosti, potom ju nemôžu spĺňať ani pravidlá {1} ⇒ {2, 4}, {2} ⇒ {1, 4} a nemusia sa teda vôbec uvažovať.

Z hore uvedeného vyplýva, že kombinácie *T* začíname rozdeľovať tak, že najskôr záver *Y* je tvorený kombináciou dĺžky 1 (položková množina *Y* je jednoprvková). Následne u tých pravidiel, ktoré dosiahnu minimálnu spoľahlivosť sa do *Y* pridá ďalšia položka z *X* atd.
Pre úplnosť uveďme, že existujú aj ďalšie postupy ako získať pravidlá.

3. Sekvenčné pravidlá

V predchádzajúcej časti sme sa zaoberali pravidlami, kde sa neuvažovalo s faktorom času. Pri riešení niektorých problémov potrebujeme nájsť asociácie medzi udalosťami, ktoré sa odohrávajú v rôznom čase. Typickým problémov je analýza správania používateľa pri prechádzaní webu, ktorá sa aplikuje na automaticky ukladané dáta v logovacích súboroch (zachytávajú informácie o IP adrese používateľovho počítača, dátume a čase prístupu, navštívenom URL a pod.). Táto metóda bola odvodená z asociačných pravidiel a je príkladom metódy zohľadňujúcej zvláštnosti Internetu.

Sekvenčné pravidlá sa úspešne aplikujú aj do ďalších oblastí, príkladom môže byť databáza porúch v telekomunikačnej sieti, respektíve databáza transakcií v banke, kde príkladom sekvenčného pravidla môže byť: Ak klient požiada o urovnanie účtu, je vysoko pravdepodobné, že v blízkom čase zatvorí všetky účty v danej banke. Sekvenčná analýza vyžaduje identifikovať klienta s časovou nálepkou každej ním vykonanej transakcie.

Niekedy presný časový údaj o udalosti nie je potrebný, stačí nám iba informácia o tom, že niektorá udalosť predchádzala udalosti inej. Príkladom takýchto sekvencií môžu byť opakované nákupy v supermarkete, kde udalosťou je jeden nákup jedného zákazníka a sekvenciou je zoznam nákupov tohto zákazníka.

V prípade analýzy správania používateľa pri prechádzaní webu udalosťou by bola návšteva konkrétnej webovej časti a sekvenciou zoznam navštívených webových častí. Používateľa by sme identifikovali IP adresou a navštívenú webovú časť URL adresou, inak povedané IP adresa by identifikovala sekvenciu a URL adresa by predstavovala udalosť v rámci tejto sekvencie. Časový údaj by v tomto prípade tiež nebol potrebný stačila by nám iba informácia o tom, že niektorá udalosť predchádzala udalosti inej.

Ako sme už spomínali sekvenčné pravidlá sú odvodené od asociačných, rozdiely preto nie sú príliš veľké [8]. *K* - sekvencia je sekvencia dĺžky *k*, t. j. obsahuje *k* stránok. Frekventovaná *k* - sekvencia je obdobou frekventovanej *k* - položkovej množiny, resp. kombinácie. Zvyčajne, frekventovaná 1 - položková množina je rovnaká ako frekventovaná 1 - sekvencia.

Na nasledujúcom príklade naznačíme rozdiely medzi algoritmom Apriori a AprioriAll, kde algoritmus AprioriAll slúži na hľadanie sekvenčných pravidiel:

D = {*S*1 = {*U*1, <*A, B, C*>}, *S*2 = {*U*2, <*A, C*>}, *S*3 = {*U*1, <*B, C, E*>}, *S*4 = {*U*3, <*A, C, D, C, E*>}}, kde *D* je databáza transakcií s časovou nálepkou, *A, B, C, D, E* sú webové časti a *U*1, *U*2, *U*3 sú používatelia. Každá transakcia je identifikovaná používateľom. Predpokladajme, že *min s* = 30%.

V tomto príklade má používateľ *U*1 aktuálne dve transakcie. Pri hľadaní sekvenčných pravidiel uvažujeme jeho sekvenciu ako aktuálne spojenie webových častí v transakciách *S*1 a *S*3, t. j. sekvencia môže pozostávať z viacerých transakcií, pričom sa nevyžadujú súvislé prístupy na stránky. Rovnako podpora sekvencie je určená nie percentom transakcií, ale percentom používateľov, ktorý majú danú sekvenciu. Sekvencia je veľká/frekventovaná ak sa

prinajmenšom nachádza v jednej sekvencii identifikovanej používateľom a spĺňa podmienku minimálnej podpory.

Prvým krokom je zoradenie transakcií podľa používateľa a s časovou nálepkou každej ním navštívenej stránky, zostávajúce kroky sú podobné ako v algoritme Apriori.

Po zoradení sme získali aktuálne sekvencie identifikované používateľom, ktoré predstavujú kompletné odkazy od jedného používateľa:

$D = \{S1 = \{U1, <A, B, C>\}, S3 = \{U1, <B, C, E>\}, S2 = \{U2, <A, C>\}, S4 = \{U3, <A, C, D, C, E>\}\}$.

Podobne ako v algoritme Apriori začíname generovaním množiny kandidátov dĺžky 1: $C1 = \{<A>, <B>, <C>, <D>, <E>\}$, z nej určíme množinu $L1 = \{<A>, <B>, <C>, <D>, <E>\}$ jednoprvkových sekvencií takých, kde každá stránka je odkazovaná prinajmenšom jedným používateľom.

Následne množiny kandidátov $C2$ vygenerujeme z $L1$ tzv. úplným spájaním, t. j. zohľadňujeme, že používateľ webu prehľadáva stránky dopredu alebo dozadu. Z tohto dôvodu algoritmus Apriori nemôže byť použitý na objavovanie znalostí z webových prístupov (web log mining), na rozdiel od algoritmu AprioriAll, ktorý spomínanú skutočnosť zohľadňuje.

Z množiny kandidátov dĺžky 2: $C2 = \{<A, B>, <A, C>, <A, D>, <A, E>, <B, A>, <B, C>, <B, D>, <B, E>, <C, A>, <C, B>, <C, D>, <C, E>, <D, A>, <D, B>, <D, C>, <D, E>, <E, A>, <E, B>, <E, C>, <E, D>\}$, určíme množinu $L2 = \{<A, B>, <A, C>, <A, D>, <A, E>, <B, C>, <B, E>, <C, B>, <C, D>, <C, E>, <D, C>, <D, E>\}$ dvojprvkových sekvencií takých, kde sa každá sekvencia prinajmenšom nachádza v jednej sekvencii identifikovanej používateľom.

Analogicky postupujeme ďalej.

$C3 = \{<A, B, C>, <A, B, D>, <A, B, E>, <A, C, B>, <A, C, D>, <A, C, E>, <A, D, B>, <A, D, C>, <A, D, E>, <A, E, B>, <A, E, C>, <A, E, D>, <B, C, A>, <B, C, E>, <B, C, D>, <B, E, A>, <B, E, C>, <B, E, D>, <C, B, A>, <C, B, E>, <C, B, D>, <C, D, A>, <C, D, B>, <C, D, E>, <C,E,A>, <C, E, B>, <C, E, D>, <D, C, A>, <D, C, B>, <D, C, E>, <D, E, A>, <D,E,C>\}$,

$L3 = \{<A,B,C>, <A,B,E>, <A,C,B>, <A,C,D>, <A,C,E>, <A,D,C>, <A,D,E>, <B,C,E>, <C,B,E>, <C,D,E>, <D,C,E>\}$,

$C4 = \{<A,B,C,E>, <A,B,C,D>, <A,B,E,C>, <A,B,E,D>, <A,C,B,E>, <A,C,B,D>, <A,C,D,B>, <A,C,D,E>, <A,C,E,B>, <A,C,E,D>, <A,D,C,B>, <A,D,C,E>, <A,D,EB>, <A,D,E,C>, <B,C,E,A>, <B,C,E,D>, <C,B,E,A>, <C,B,E,D>, <C,D,E,A>, <C,D,E,B>, <D,C,E,A>, <D,C,E,B>\}$,

$L4 = \{<A,B,C,E>, <A,C,B,E>, <A,C,D,E>, <A,D,C,E>\}$,

Z tejto množiny štvorprvkových sekvencií už nie je možné vygenerovať žiadneho päťprvkového kandidáta.

$C5 = \emptyset$.

4. Záver

Prostredníctvom sekvenčnej analýzy môžeme odhaliť zaujímavé vzorce správania sa používateľov rôznych portálov/systémov. Dokážeme popísať, ktoré webové časti jednotliví používatelia navštívili za sledované obdobie, a v akom poradí sa návštevy uskutočnili.

V nasledujúcich odstavcoch uvádzame na základe vlastných skúseností možné aplikácie spomínanej metódy a navrhujeme postupy na riešenie problémov z oblasti objavovania znalostí na základe používania webu.

Jednou z možných aplikácii je optimalizácia webu na základe nájdených znalostí. Časť nájdených pravidiel nám môže odhaliť nedostatky portálu vyplývajúce z jeho používania. Odstránením identifikovaných nedostatkov môžeme prispieť k efektívnejšiemu používaniu portálu. Druhá časť pravidiel nám odhaľuje vzorce správania sa používateľov pri prechádzaní

portálu na základe, ktorých môžeme preskupiť odkazy a pod. Nájdené pravidlá predstavujú vzorce správania používateľov portálu, prostredníctvom ktorých môžeme predikovať návštevy rôznych webových častí, respektíve optimalizovať portál, napr. objavením zavádzajúcich odkazov a pod. Dôležité si je uvedomiť, že niektoré nájdené pravidlá môžu mať sezónny charakter. Homogénne skupiny používateľov pri prechádzaní portálu sa budú pravdepodobne správať inak v rôznych časových obdobiach. Z tohto dôvodu je potrebné pri analýze zohľadňovať faktor sezónnosti a analýzu realizovať v jednotlivých časových obdobiach.

Navrhujeme nasledovný postup:

1. získavanie dát – definovanie sledovaných premenných do logovacieho súboru z pohľadu získania potrebných dát (IP adresa, dátum a čas prístupu, URL adresa, protokol, webový prehliadač, webová lokalita, rozlíšenie obrazovky a pod.),
2. príprava dát – vytvorenie dátových matíc z logovacieho súboru (informácie o prístupoch) a mapy webu (informácie o obsahu webu), identifikovanie IP adries (prehľadávacie stroje (google, yahoo a pod.), IP adresy smerovačov a pod.),
3. analýza dát - zohľadňovanie sezónnosti realizáciou analýzy v jednotlivých časových obdobiach,
4. optimalizácia webu – odstránenie identifikovaných nedostatkov vyplývajúcich z používania portálu a predikovanie správania sa používateľov.

Ďalšou možnou aplikáciou je použitie nájdených znalostí ako podkladu pre tvorbu adaptívnych hypermediálnych systémov. V tomto prípade by sme si nevystačili s dátami z logovacích súborov a výstupom sekvenčnej analýzy. Aj keď tieto dáta spolu s nájdenými pravidlami by boli pre nás kľúčové. Ešte predtým ako by sme aplikovali samotnú sekvenčnú analýzu je nutné segmentovať používateľov systému, t. j. rozdeliť používateľov do viacerých zrozumiteľných podskupín a analyzovať štruktúru každej z nich, čo umožní hľadať špecifické sekvenčné pravidlá pre každú skupinu osobitne. Objavením významných znakov v správaní súčasných používateľov, potom môžeme výrazne prispieť napríklad k dobrým predikciám správania nových používateľov systému alebo k zvýšeniu efektivity používania systému. Aké dáta je potrebné získať od používateľov a aké metódy použiť na nájdenie potrebných znalostí záleží od viacerých faktorov. Rozhodnutie na základe akých dát segmentovať používateľov závisí od zamerania systému, podobne ako výber metódy na segmentáciu používateľov závisí od charakteru dát. Nevyhnutnosťou pri analýze a interpretácii je dobre poznať vecnú povahu dát, aby sme nájdené znalosti mohli použiť ako podklad pre tvorbu adaptívnych hypermediálnych systémov. Napríklad, v prípade elektronického obchodu vieme získať dáta od zákazníkov pri registrácii, pričom by nás pravdepodobne zaujímali informácie zamerané na socio - ekonomické charakteristiky, ako plat, vek, pohlavie, profesia a pod. Aplikovaním napríklad zhlukovej analýzy môžeme rozdeliť svojich zákazníkov do viacerých zrozumiteľných podskupín a analyzovať štruktúru každej z nich, čo umožní rozvíjať špecifické stratégie pre každú skupinu osobitne. Použitie takýchto znalostí pri tvorbe adaptívnych hypermediálnych systémov by bolo možné za podmienky, že by sa nejednalo o portál/systém s anonymným prístupom, vzhľadom na to, že by nebolo možné identifikovať homogénne skupiny používateľov portálu, pre ktoré by bolo potrebné hľadať špecifické vzorce správania. Z tohto pohľadu použitie IP adresy na identifikáciu používateľov sa javí ako nie najvhodnejšie riešenie. Tento problém by v budúcnosti mohol vyriešiť prechod na IPv6, vďaka čomu by sme vedeli ľahšie identifikovať jednotlivých používateľov. V súčasnosti sa javí ako najvýhodnejšie analyzovať prístupy do systémov, kde sa vyžaduje autentifikácia.

Navrhujeme nasledovný postup:

1. definovanie zamerania systému (z hľadiska potrieb používateľov),
2. získavanie dát – definovanie sledovaných premenných do logovacieho súboru a formulára z pohľadu získania potrebných dát,
3. príprava dát – vytvorenie dátových matíc,
4. identifikovanie dostatočne veľkých homogénnych skupín používateľov (pokiaľ ide o úroveň potrieb),
5. určenie profilu segmentov (v zmysle jeho odlišných postojov v správaní, v socio – ekonomických, v psycho – grafických, v demografických charakteristikách a pod.),
6. nájdenie vzorcov správania sa používateľov v jednotlivých segmentoch,
7. predikovanie správania sa používateľov v jednotlivých segmentoch (pričom zohľadňujeme špecifické potreby cieľového segmentu, resp. segmentov).

5. Literatúra

[1] ZAIANE, O. – HAN, J. 1998. WebML: Querzing the World-Wide Web for resources and knowledge. In: Proc. Int. Workshop on Web Information and Data Management. WIDM'98, 1998, s. 9-12.

[2] BERKA, P. 2003. Dobývání znalostí z databází. Praha: Academia, 2003. 392 s. ISBN 80-200-1062-9.

[3] SRIVASTAVA, J. - COOLEY, R. - DESHPANDE, M. - TAN, P. 2000. Web usage mining: discovery and applications of web usage patterns from web data. In: SIGKDD Explorations, č. 1, 2000, s. 12 - 23.

[4] AGRAWAL, R. – IMIELINSKI, T. – SWAMI, A. 1993. Mining association rules between sets of items in large databases. In: ACM SIGMOD Conference, 1993, s. 207–216.

[5] HEMERT, J. – BALDOCK, R. 2007. Mining Spatial Gene Expression Data for Association Rules. Lecture Notes in Computer Science, č. 4414, 2007, s. 66-76.

[6] AGRAWAL, R. - SRIKANT, R. 1994. Fast algorithms for mining association rules in large databases. In: VLDB Conference, 1994, s. 487–499.

[7] HIPP, J. – GUNTZER, U. – NAKHAEIZADEH, G. 2000. Algorithms for Association Rule Mining - A General Survey and Comparison. SIGKDD Explorations, č. 2, 2000, s. 58-64.

[8] WANG TONG - HE PI-LIAN. 2005. Web Log Mining by an Improved AprioriAll Algorithm. In: Engineering and Technology, č. 4, 2005, s. 97-100.

Adresa autora:

Michal Munk, RNDr., PhD.
Fakulta prírodných vied UKF
Katedra informatiky
Tr. A. Hlinku 1
949 74 Nitra
mmunk@ukf.sk

Prípadová štúdia analýzy spoľahlivosti dotazníka
Case Study of Reliability Analysis of Questionnaire

Michal Munk, Ján Záhorec

Abstract: The paper provides case study of reliability/item analysis of questionnaire utilization. The aim of the questionnaire was to verify possibilities to affect students' attitude towards less favourite subjects by means of pedagogical intervention of electronic teaching materials and interactive animations Principles of Geometry Optics in educational process at particular schools.

Key words: Reliability, reliability/item analysis, Cronbach's alpha coefficient.

Kľúčové slová: Reliabilita, analýza spoľahlivosti/položiek, Cronbachov koeficient alfa.

1. Úvod

Využívanie informačných a komunikačných technológií a multimédiami podporovaného vyučovania na zvyšovanie kvality vzdelávania má vzostupnú tendenciu, čo súvisí s prenikaním nových technológií do všetkých sfér praktického života. Popri zavádzaní moderných multimediálnych výučbových prostriedkov do vyučovania menšia pozornosť je venovaná možnostiam ovplyvňovania postojov študentov k študijným predmetom a zvyšovaniu učebnej motivácie prostredníctvom zaraďovania uvedených médií do výučby jednotlivých predmetov. Túto skutočnosť považujeme za veľmi dôležitú hlavne vo vzťahu k tzv. menej obľúbeným alebo neobľúbeným predmetom. Nakoľko neobľúbenosť prírodovedných disciplín u študentov pripisujeme na margo nedostatku príťažlivých motivujúcich výučbových materiálov, rozhodli sme sa pre potreby nášho výskumu vyvinúť multimediálne učebné materiály pod názvom *Základy geometrickej optiky* na podporu výučby základov geometrickej optiky. Geometrickú optiku sme si vybrali vzhľadom na skutočnosť, že táto téma je súčasťou učebných osnov fyziky na všetkých typoch stredných škôl a navyše sa vyučuje na základných školách. Okrem toho spracovanie tejto témy nebýva pre študentov príliš príťažlivé vďaka značnej abstraktnosti dôležitých pojmov nutných k pochopeniu problematiky. Niektoré pojmy navyše prakticky nie je možné vysvetliť pomocou klasickej školskej učebnice, ktorá nedisponuje interaktívnymi animáciami, ktoré môžu k vysvetleniu niektorých javov a pojmov prispieť zásadným spôsobom.

Úlohou multimediálneho učebného materiálu *Základy geometrickej optiky* je predovšetkým podporiť a zatraktívniť klasické prezenčné formy výučby fyziky, prispieť ku skvalitneniu individuálnej práce a samoštúdia učiacich sa. Jeho jednotlivé časti môže učiteľ využívať v rámci multimédiami podporovaného vyučovania ale takisto ho môžu používať aj samotní študenti, či už ako podporný ilustračný študijný materiál v rámci samostatnej práce na vyučovaní alebo v rámci domácej prípravy na vyučovanie.

Na overenie stanovenej hypotézy výskumu, v ktorej *predpokladáme, že vyučovanie podporované elektronickými výučbovými prostriedkami prispieva k znižovaniu negatívnych postojov k vyučovacím predmetom, konkrétne k vyučovaciemu predmetu fyzika*, sme sa rozhodli použiť pedagogický experiment. V rámci pedagogického experimentu boli sledované a vyhodnocované možnosti ovplyvňovania negatívnych postojov a vzťahov študentov k fyzike prostredníctvom pedagogickej intervencie nami vytvorených multimediálnych učebných materiálov *Základy geometrickej optiky* do vyučovacieho procesu.

Pedagogický experiment bol vykonaný v rámci predmetu fyzika na štvorročnom a osemročnom gymnáziu na Golianovej ulici v Nitre. Výskumnú vzorku tvorili študenti 4. a 8.

(oktáva) ročníka, čo predstavuje vekovú úroveň 17-19 rokov. Celkový výskumný súbor bol tvorený 82 študentmi zapojenými do výskumu, z ktorých bolo 58,5 % chlapcov a 41,5 % dievčat. Pedagogická intervencia nami vyvinutých materiálov do vyučovacieho procesu bola realizovaná jednak počas klasických vyučovacích hodín a jednak počas seminárov z fyziky.

Výskum prebiehal v nasledujúcich fázach:

1. Príprava experimentálnych učebných materiálov. Konštrukcia dotazníkov experimentu.
2. **Posúdenie kvality výskumných nástrojov (analýza spoľahlivosti/položiek).**
3. Vytvorenie kontrolnej a experimentálnej skupiny.
4. Administrovanie dotazníkov.
5. Realizácia experimentu.
6. Readministrovanie dotazníkov.
7. Evaluácia vytvorených e-produktov *Základy geometrickej optiky.*
8. Exploračná analýza dát.
9. Overenie validity použitých štatistických metód.
10. Analýza dát.
11. Interpretácia výsledkov výskumu.

Proces merania je predpokladom získania dát. Aby meranie bolo kvalitné, musí byť meracia procedúra objektívna, reliabilná a validná. Cieľom druhej fázy realizovaného pedagogického experimentu bolo posúdenie reliability administrovaného dotazníka *Zisťovanie vzťahu a postojov študentov k vyučovaciemu predmetu fyzika* a identifikovanie podozrivých položiek. Pri skúmaní ďalších problémov, musíme zaručiť, že dokážeme odhadnúť vplyv kvality meracej procedúry na naše výsledky experimentu.

2. Analýza spoľahlivosti/položiek

Analýza spoľahlivosti/položiek patrí medzi viacrozmerné prieskumné techniky a slúži k posúdeniu kvality – spoľahlivosti meracej procedúry, napríklad škály dotazníka a k identifikovaniu podozrivých položiek. K priamym odhadom spoľahlivosti patrí Cronbachov koeficient alfa

$$\hat{\alpha} = \frac{m}{m-1} \cdot \left(1 - \frac{\sum s_j^2}{s^2} \right), \tag{1}$$

kde m je počet položiek dotazníku, s^2 je rozptyl škály dotazníku, s_j^2 je rozptyl škály j - tej položky dotazníku [1].

Odhad reliability môžeme dostať aj z priemerného korelačného koeficientu $\bar{r}$ jednotlivých položiek. Nazývame ho štandardizovaný Cronbachov koeficient alfa

$$\overline{\alpha} = \frac{m\bar{r}}{1 + (m-1)\bar{r}}, \tag{2}$$

kde *m* je počet položiek [1].

Štandardizovaný Cronbachov koeficient alfa dostaneme aj z predchádzajúceho vzťahu, ak sme všetky merania dopredu štandardizovali, t. j. od každej hodnoty premennej sa odpočíta jej priemer a vydelí sa jej smerodajnou odchýlkou.

Ak sú obidva odhady príliš odlišné, indikuje to, že jednotlivé položky nemajú rovnakú variabilitu [2].

3. Dotazník *Zisťovanie vzťahu a postojov študentov k vyučovaciemu predmetu fyzika*

Názory respondentov v položkách dotazníka sú zaznamenané na sedemstupňovej škále, resp. na päťstupňovej škále. Vyššia miera nesúhlasu s predloženým tvrdením (otázkou) je označená nižšou hodnotou, úplný nesúhlas je označený stupňom 1, vyššia miera súhlasu s predloženým tvrdením (otázkou) je označená vyššou hodnotou, úplný súhlas je označený stupňom 7 (resp. 5). U každého študenta bola pri spomínaných položkách v rámci pretestu ako aj v rámci posttestu zaznamenaná hodnota škály podľa toho, akú mieru svojho súhlasu alebo nesúhlasu s jednotlivými tvrdeniami vyznačil. Okrem toho sme u každého študenta zaznamenali známku z fyziky na konci 3. ročníka, ktorá predstavovala taktiež ordinálnu premennú.

4. Analýza spoľahlivosti dotazníka

Na analýzu sme použili techniky a metódy na posúdenie spoľahlivosti dotazníka a identifikovanie jeho podozrivých položiek.

Z korelačnej matice uvedenej v tabuľke 1 môžeme identifikovať podozrivé položky dotazníka.

Tabuľka 1: Korelačná matica položiek dotazníka (pretest)

	1PRE	2PRE	3PRE	4PRE	5PRE	Známka z fyziky
1PRE	1,000000	0,601537	0,245188	0,447306	0,370703	-0,335337
2PRE	0,601537	1,000000	0,099246	0,371584	0,485987	-0,167226
3PRE	0,245188	0,099246	1,000000	0,314011	0,196405	-0,198659
4PRE	0,447306	0,371584	0,314011	1,000000	0,282400	-0,335042
5PRE	0,370703	0,485987	0,196405	0,282400	1,000000	-0,033369
Známka z fyziky	-0,335337	-0,167226	-0,198659	-0,335042	-0,033369	1,000000

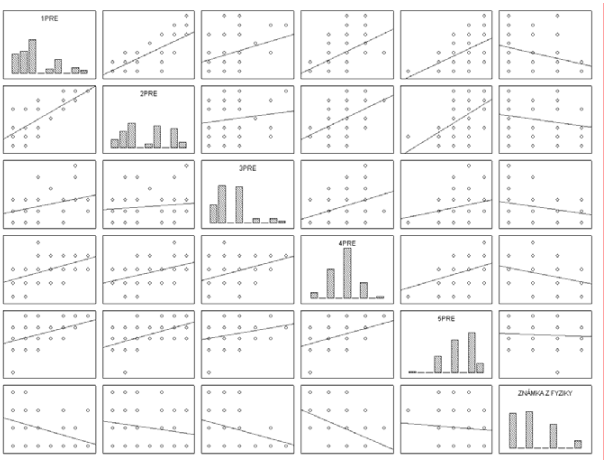
Zvýraznené korelačné koeficienty sú štatisticky významné na hladine významnosti 0,05.

Z korelačnej matice tabuľky 1 vidíme, že medzi väčšinou položiek sú korelácie štatisticky významné, čo znamená, že medzi týmito položkami existuje určitá miera vzájomnej závislosti. Čím viac sa korelačný koeficient približuje k hodnote 1, tím je priamoúmerná závislosť silnejšia.

Výnimkou sú položky 2 a 3 medzi ktorými nie je korelácia štatisticky významná, z čoho môžeme usúdiť, že hodnoty sa menia nezávisle. Tieto položky sa na základe týchto výsledkov javia ako podozrivé.

Medzi položkou *Známka z fyziky* a všetkými ostatnými položkami je negatívna korelácia, t.j. hodnoty sa menia spoločne, ale v opačnom smere (kým hodnoty jednej premennej klesajú druhej premennej rastú).

Korelačná matica jednotlivých položiek pretestu tabelovaná v tabuľke 1 je vizualizovaná v grafe 1.



Obrázok 1: Maticový graf – vizualizácia korelačnej matice

Tabuľka 2: Súhrnné štatistiky dotazníka (pretest)

Počet položiek dotazníka			5
Počet platn. prípadov			73
Priemer	16,534246575	**Súčet**	1207,0000000
Smerod. odch.	4,672859557	**Rozptyl**	21,835616438
Minimum	6,000000000	**Maximum**	28,000000000
Cronbachova alfa	0,767788059	**Standardiz. alfa**	0,782071554
Priemerná korelácia medzi položkami			0,430876323

Hodnota koeficientu reliability 0,77 (77 %) vyjadruje podiel variability súčtu škály položiek k celkovej variabilite dotazníka. Obidva odhady (Cronbachova alfa a štandardizovaná alfa) nie sú príliš odlišné, t. j. jednotlivé položky majú rovnakú variabilitu (Tabuľka 2).

Dotazník môžeme považovať za spoľahlivý, avšak nízka priemerná korelácia medzi položkami naznačuje, že po odstránení niektorých položiek by sme mohli spoľahlivosť dotazníka zvýšiť.

Z tabuľky 3 vidíme, že všetky položky pretestu korelujú s celkovým skóre škály a po odstránení klesol koeficient reliability. U tretej položky sledujeme opačný stav, v tomto prípade koeficient reliability vzrástol.

Po odstránení 3. položky sa zvýšil koeficient reliability - Cronbachova alfa z 0,77 na 0,79. Z toho vyplýva, že 3. položka znižuje celkovú spoľahlivosť dotazníka, preto ju podrobíme hlbšej kvalitatívnej analýze.

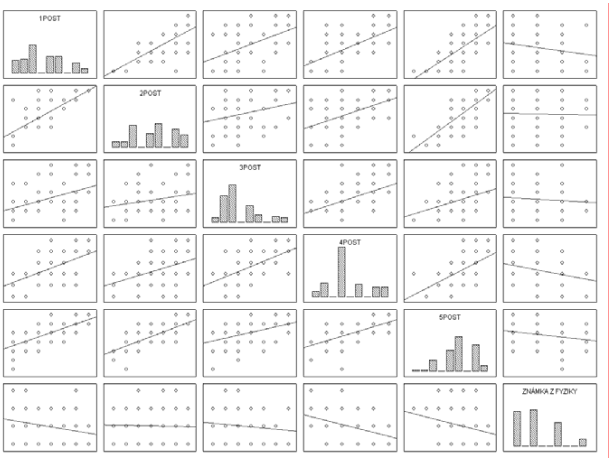
Tabuľka 3: Štatistiky dotazníka (pretest) po odstránení príslušnej položky

	Priem. po odstr.	Rozptyl po ods.	Sm.odch. po ods.	Pol.-Celk. korel.	Alfa po odstr.
1PRE	13,54795	11,26140	3,355801	0,731071	0,646448
2PRE	12,58904	10,79002	3,284816	0,657158	0,692267
3PRE	14,09589	17,67574	4,204252	0,314200	0,789265
4PRE	13,75342	17,03509	4,127359	0,570327	0,735052
5PRE	12,15069	16,15538	4,019375	0,577968	0,723691

Analýza spoľahlivosti posttetu nám iba potvrdila výsledky analýzy spoľahlivosti pretestu.

Tabuľka 4: Korelačná matica položiek dotazníka (posttest)

	1POST	2POST	3POST	4POST	5POST	Známka z fyziky
1POST	1,000000	0,630168	0,355395	0,391970	0,535228	-0,175595
2POST	0,630168	1,000000	0,136654	0,292375	0,597861	-0,027205
3POST	0,355395	0,136654	1,000000	0,392879	0,273538	-0,092467
4POST	0,391970	0,292375	0,392879	1,000000	0,433043	-0,215245
5POST	0,535228	0,597861	0,273538	0,433043	1,000000	-0,177702
Známka z fyziky	-0,175595	-0,027205	-0,092467	-0,215245	-0,177702	1,000000



Obrázok 2: Maticový graf – vizualizácia korelačnej matice

Tabuľka 5: Súhrnné štatistiky dotazníka (posttest)

Počet položiek dotazníka			5
Počet platn. prípadov			73
Priemer	19,520547945	**Súčet**	1425,0000000
Smerod. odch.	5,671834871	**Rozptyl**	32,169710807
Minimum	8,000000000	**Maximum**	34,000000000
Cronbachová alfa	0,840292399	**Standardiz. alfa**	0,843874534
Priemerná korelácia medzi položkami			0,536805848

Tabuľka 6: Štatistiky dotazníka (posttest) po odstránení príslušnej položky

	Priem. po odstr.	Rozptyl po ods.	Sm.odch. po ods.	Pol.-Celk. korel.	Alfa po odstr.
1POST	15,95890	18,20379	4,266590	0,781328	0,766232
2POST	15,05479	19,06549	4,366405	0,680823	0,799711
3POST	16,23288	23,68550	4,866775	0,451928	0,856600
4POST	16,16438	22,63051	4,757154	0,623301	0,814828
5POST	14,67123	22,00150	4,690576	0,735138	0,790543

5. Analýza podozrivých položiek dotazníka

Na základe výsledkov výskumu sme realizovali podrobnejšiu analýzu podozrivej položky číslo 3, t. j. položky ktorá po odstránení zvyšovala spoľahlivosť dotazníka.

Položka č. 3: Fyzika patrí medzi predmety... (a) veľmi nenáročné, b) nenáročné, c) skôr nenáročné, d) ani náročné, ani nenáročné, e) skôr náročné, f) náročné, g) veľmi náročné).

Výrokom položeným v tretej položke sme chceli získať názor študentov na náročnosť vyučovacieho predmetu fyzika na gymnáziu. Túto položku dotazníka sme pokladali za jednu z položiek, ku ktorej respondenti vyjadria svoje stanovisko s jednoznačným výsledkom náročnosti fyzikálneho obsahu vzdelávania stanoveného učebnými osnovami. Prostredníctvom štatistických metód sme však prišli k zaujímavým zisteniam. Meranie pomocou škály ukázalo, že patrila medzi položky, u ktorých sa vyskytli najdivergujúcejšie odpovede a po jej odstránení sa najviac zvýšil koeficient reliability dotazníka. Táto položka znižovala spoľahlivosť celého dotazníka z dôvodu, že aj študenti, ktorí na škále ostatných položiek vyjadrili súhlas s tým, že fyzika je pre nich skôr obľúbená, zaujímavá a významná, odpovedali na túto položku z výsledkom náročnosti fyziky. Takmer 2/3 respondentov z celého dátového súboru pedagogického experimentu volila možnosť *e* (skôr náročné), *f* (náročné), alebo *g* (veľmi náročné).

6. Záver

Analýzou spoľahlivosti/položiek môžeme zvýšiť reliabilitu dotazníka, respektíve môžeme zabrániť použitiu nekvalitného dotazníka, prostredníctvom ktorého získané dáta by nemali žiadnu výpovednú hodnotu, bez ohľadu na to, akú pokročilú metódu na ich ďalšie spracovanie použijeme.

Aplikáciou prezentovanej analýzy sme získali spoľahlivé experimentálne dáta prostredníctvom ktorých sme mohli overiť hypotézu, že vyučovanie podporované elektronickými výučbovými prostriedkami prispieva k znižovaniu negatívnych postojov k vyučovacím predmetom, konkrétne k vyučovaciemu predmetu fyzika v danej vekovej kategórii študentov.

7. Literatúra

[1] STATSOFT, INC. 1999. Electronic Statistics Textbook. Tulsa, OK: StatSoft, 1999. WEB: http://www.statsoft.com/textbook/stathome.html

[2] MUNK, M. - KLOCOKOVÁ, D. - LANČARIČ, D. – ČERVEŇANSKÁ, M. 2008. Tvorba, správa a analýza e-kurzov. Nitra: UKF, 2008. 160 s. ISBN 978-80-8094-118-5.

Adresa autorov:

Michal Munk, RNDr., PhD.
Fakulta prírodných vied UKF
Tr. A. Hlinku 1, 949 74 Nitra
mmunk@ukf.sk

Ján Záhorec, PaedDr., PhD.
Pedagogická fakulta UKF
Drážovská cesta 4, 949 74 Nitra
jzahorec@ukf.sk

Monitoring prediktivního modelu - indexy stability populace
Predictive Model Monitoring - Population Stability Indices

Martin Řezáč

Abstract: Predictive models help streamline the decision-making process by improving the quality and consistency of decisions being made. In order to achieve maximum ongoing effectiveness and maintain profitability, front-end reports should be put in place to track model stability through the model's entire life cycle. Some compendious indices, like Population Stability Index (PSI), Gini index and Kolmogorov-Smirnov statistics (K-S index), can be used. In cases where these indices identify model misbehaviour, a characteristic analysis should be applied to the current and development populations to identify which characteristics (attributes) could be the causes of that misbehaviour. This paper is focused on the compendious indices. Some results on real financial data are included.

Key words: Population Stability Index (PSI), predictive model monitoring, Gini index, K-S index.

Klíčová slova: Index stability populace (PSI), monitoring prediktivního modelu, Giniho index, K-S index.

1. Úvod

Není překvapivé, že prediktivní modely se ve statistickém slova smyslu chovají nejlépe na vývojovém vzorku dat. Výstupy těchto modelů, např. skóre nebo rating klienta, jsou počítány pomocí jistých vzorců, jejichž koeficienty příslušející nezávislým proměnným (prediktorům) jsou odvozeny na datech vývojového vzorku. Posun distribuce výstupu daného modelu je pak zapříčiněn právě změnou vstupních hodnot modelu, tj. prediktorů, v průběhu času. V podstatě ihned (alespoň většinou) po nasazení prediktivního modelu do praxe dochází k jistému poklesu jeho prediktivní síly, který je způsoben určitou změnou vstupních hodnot modelu. Zásadní je v praxi nastavení takových procesů, které odhalí, že se tak děje, proč se tak děje a jak vážný problém to ve svých důsledcích znamená.

Faktorů způsobujících posun v distribuci prediktorů, a následně posun v distribuci výstupu prediktivního modelu, je několik:

- Přirozený posun v datech/změna demografické struktury dat
- Databázové chyby
- Změna datového zdroje
- Změna definice/formátu vstupních dat
- Změna datového univerza
- Ostatní

Typickým příkladem prvního uvedeného důvodu je příjem klienta (všeobecným trendem je růst příjmu populace). Změnou definice/formátu vstupních dat je myšlena například situace, kdy je rozšířen číselník hodnot, kterých může vstupní proměnná nabývat. Změnou datového univerza je myšlen případ kdy je vyvinutý prediktivní model použit např.pro odlišný/nový segment portfolia nebo odlišný/nový produkt.

2. Prediktivní síla modelu

Základním sledovaným kritériem daného modelu je jeho stabilita prediktivní schopnosti v čase. Nejčastěji užívanými mírami této schopnosti jsou Giniho index a K-S index (Kolmogorovova-Smirnovova statistika).

Předpokládejme, že máme posoudit sílu modelu předikujícího klientovu schopnost splácet své finanční závazky. Klienty, kteří splní jistou definici vyjadřující tuto schopnost, nazveme dobrými klienty, ostatní klienty špatnými. Zaveďme následující označení.

$$D_K = \begin{cases} 1, & je-li\ klient\ dobrý \\ 0, & je-liklient\ špatný \end{cases}$$

Výstupem modelu je typicky skóre s, což je číslo z intervalu [0,1] kvantifikující pravděpodobnost schopnosti splácet. Distribuční funkce skóre dobrých, resp. špatných, klientů jsou dány vztahem

$$F_{GOOD}(a) = P\left(s \le a \middle| D_K = 1\right),$$
$$F_{BAD}(a) = P\left(s \le a \middle| D_K = 0\right),\ a \in [0,1].$$

Jejich empirické podoby jsou dány vztahy

$$F_{n.GOOD}(a) = \frac{1}{n}\sum_{i=1}^{n} I\left(s_i \le a \wedge D_K = 1\right),$$
$$F_{m.BAD}(a) = \frac{1}{m}\sum_{i=1}^{m} I\left(s_i \le a \wedge D_K = 0\right),\ a \in [0,1],$$

Giniho index lze, vzhledem k předchozím označením, vypočítat například pomocí vztahu

$$Gini = 1 - \sum_{k=2}^{N}\left(F_{m.BAD_k} - F_{m.BAD_{k-1}}\right)\cdot\left(F_{n.GOOD_k} + F_{n.GOOD_{k-1}}\right),$$

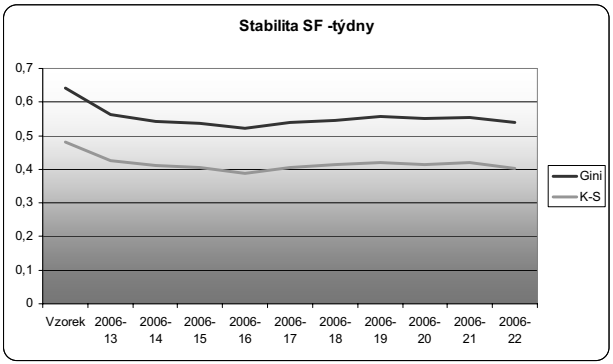
kde $F_{m.BAD_k}$, respektive $F_{n.GOOD_k}$, je k-tá hodnota vektoru empirické distribuční funkce špatných, resp. dobrých, klientů. N je dálka tohoto vektoru. Více viz [4]. Kolmogorovova-Smirnovova statistika (K-S index) je definována jako

$$KS = \sup_{a\in[0,1]}\left|F_{BAD}(a) - F_{GOOD}(a)\right|.$$

Z definice přímo plyne, že její hodnoty leží v intervalu [0,1]. Její odhad lze snadno spočítat pomocí vztahu

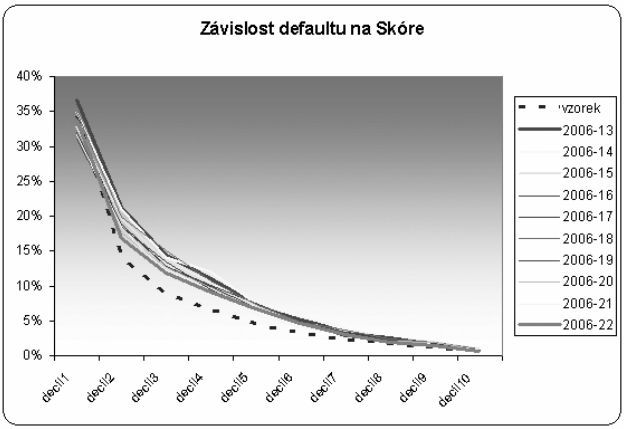
$$KS = \max_{a\in[0,1]}\left|F_{m,BAD}(a) - F_{n,GOOD}(a)\right|.$$

Na obrázku 1 je uveden příklad průběhu Giniho a K-S indexu měřeného týdně od okamžiku nasazení příslušné skóringového modelu do reálného procesu. Jako první hodnota je vynesen Giniho a K-S index na vývojovém vzorku. U obou indexů je patrný pokles oproti hodnotám na tomto vzorku a následné relativně stabilní chování.



Obrázek 1: Stabilita Giniho indexu a K-S indexu . Zdroj: vlastní konstrukce.

Na obrázku 2 je zobrazena závislost default rate, tj. relativního zastoupení nesplácejících (špatných) klientů, na skóre. Konkrétně je vynesen default rate v jednotlivých intervalech skóre, které byly určeny pomocí hodnot decilů skóre na vývojovém vzorku. Opět je zobrazena situace pro vývojový vzorek a pro následujících 10 týdnů. Na rozdíl od předchozího obrázku se jeví situace o něco dramatičtěji. Nicméně stále lze chování modelu považovat za stabilní.



Obrázek 2: Stabilita závislosti defaultu na skóre. Zdroj: vlastní konstrukce.

3. Index stability populace

Mimo monitorování prediktivní síly modelu je třeba sledovat také jak se v čase vyvíjí vstupy modelu. To má totiž za následek změnu v distribuci výstupu modelu. Předpokládejme opět, že výstupem modelu je odhad klientovi schopnosti splácet své finanční závazky, který je reprezentován pomocí skóre. Posun v distribuci skóre pak může zapříčinit neefektivnost v rozhodovacích procesech, které jsou na skóre navázány.

Zmíněný posun je ovšem třeba nějak kvantifikovat. Cílem je získat nějakou souhrnný index popisující rozdílnost dvou zkoumaných distribucí skóre. Přesněji řečeno máme provést statistický test shody daných distribucí. Předpokládejme, že máme k dispozici hodnoty decilů skóre z vývojového vzorku případně nějaké jiné pevné hranice na škále skóre. Dále jsou známy počty klientů, jejichž skóre přísluší do intervalu definovaného těmito hranicemi a to jednak pro vývojový vzorek a následně pro nějaké aktuální časové období. Nejjednodušší možností provedení testu shody je použití χ^2 statistiky. Ta je dána vztahem

$$\chi^2 = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i},$$

kde O_i (E_i) je procentní zastoupení i-tého intervalu skóre pro aktuální období (vývojový vzorek), r je počet intervalů skóre (typicky $r = 10$). S využitím tabelovaných kritických hodnot pak lze snadno provést test o shodě daných distribucí skóre.

Jinou možností je využití indexu stability populace (PSI), viz [2], [3]. Ten je definován vztahem

$$PSI = \sum_{i=1}^{r} (O_i - E_i) \ln\left(\frac{O_i}{E_i}\right),$$

kde O_i, E_i a r má stejný význam jako v případě χ^2 statistiky. Jako kritická hodnota pro test shody je v literatuře, např. [1], je uváděna hodnota 0,1 a 0,25 s následujícím významem:

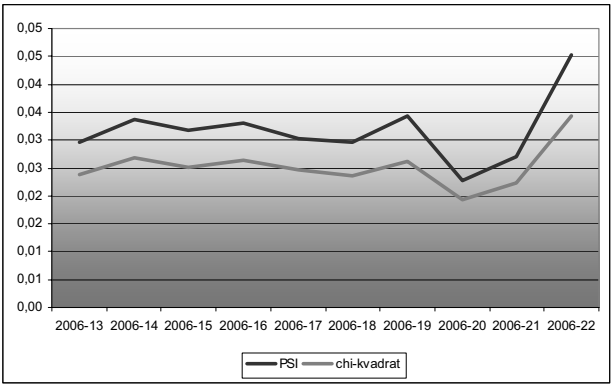
- $PSI \leq 0{,}1$ značí žádný nebo jen velmi malý rozdíl daných distribucí skóre.
- $0{,}1 < PSI \leq 0{,}25$ znamená, že došlo k nějakému posunu distribuce, nicméně nikterak významnému.
- $PSI > 0{,}25$ signalizuje významný posun v distribuci skóre, tj. zamítáme hypotézu o shodě daných distribucí.

V následující tabulce 1 je uveden příklad výpočtu PSI. Pro lepší orientaci jsou sloupce označeny pomocí čísel, navíc jsou naznačeny i matematické operace, které se v příslušném sloupci používají. Hodnota indexu PSI je pak součet hodnot uvedených v posledním sloupci.

Tabulka 1: Výpočet PSI, Zdroj: vlastní konstrukce.

	výv. vzorek [1]	týden1 [2]	[3]=[2] -[1]	[4]=[2]/[1]	[5]=ln[4]	[6]=[3]*[5]
skóre_1	10,00%	5,63%	-0,044	0,563	-0,574	0,025
skóre_2	10,00%	11,21%	0,012	1,121	0,114	0,001
skóre_3	10,00%	11,00%	0,010	1,100	0,095	0,001
skóre_4	10,00%	10,97%	0,010	1,097	0,092	0,001
skóre_5	10,00%	10,31%	0,003	1,031	0,031	0,000
skóre_6	10,00%	10,12%	0,001	1,012	0,012	0,000
skóre_7	10,01%	9,62%	-0,004	0,961	-0,039	0,000
skóre_8	10,00%	9,89%	-0,001	0,989	-0,011	0,000
skóre_9	10,00%	10,31%	0,003	1,031	0,030	0,000
skóre_10	10,00%	10,94%	0,009	1,095	0,091	0,001
					PSI	0,030

Na stejných datech jako v případě obrázků 1 a 2 byly spočteny hodnoty PSI a χ^2 statistiky. Jak je z obrázku 3 patrné, ani jedna z těchto charakteristik se ani zdaleka neblíží příslušných kritickým hodnotám. Jediné co lze vysledovat je mírný nárůst v posledním sledovaném týdnu.



Obrázek 3: Hodnoty PSI a χ^2 statistiky . Zdroj: vlastní konstrukce.

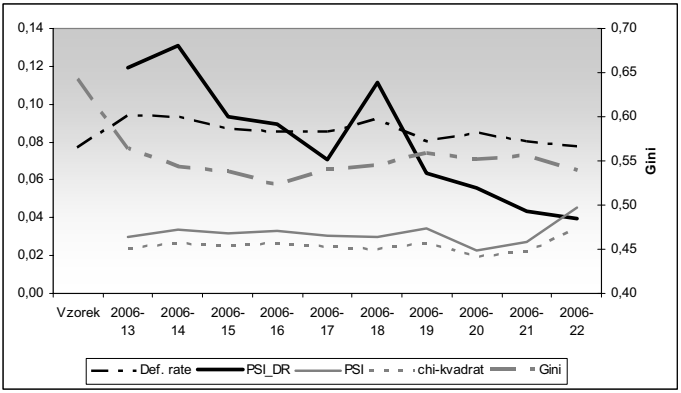
Uvažujme nyní modifikovaný index stability populace daný vztahem

$$PSI_{DR} = \sum_{i=1}^{r} (DR2_i - DR1_i) \ln\left(\frac{DR2_i}{DR1_i}\right),$$

kde $DR2_i$, resp. $DR1_i$, značí default rate v i-tém intervalu skóre na aktuálně sledovaném období, resp. na vývojovém vzorku. Takto upravený index spojuje vlastnosti indexu PSI (reaguje na změnu populace) a Giniho indexu (měří prediktivní sílu modelu) a současně silně reaguje na změnu hladiny default rate. V případě kdy se populace dramaticky nemění, tj. PSI je blízký nule, a prediktivní síla modelu je stabilní, je stále možné, že se výrazně změní hladina default rate. Zatímco Giniho index a PSI na tuto změnu v podstatě nereagují, index PSI_{DR} takovou změnu zachytí. Následující tabulka 2 obsahuje hodnoty modifikovaného PSI (stejná data jako u obr. 1,2,3). Pro srovnání jsou ještě uvedeny hodnoty default rate, Giniho indexu, PSI a χ^2 statistiky. Vše je také zobrazeno na obrázku 4.

Tabulka 2: Modifikovaný PSI, Zdroj: vlastní konstrukce.

	Def. rate	Gini	PSI_DR	PSI	chi-kvadrat
Vzorek	7,69%	0,643			
2006-13	9,38%	0,564	0,120	0,030	0,024
2006-14	9,35%	0,542	0,131	0,034	0,027
2006-15	8,70%	0,537	0,093	0,032	0,025
2006-16	8,57%	0,523	0,089	0,033	0,026
2006-17	8,59%	0,540	0,071	0,030	0,025
2006-18	9,19%	0,544	0,111	0,030	0,024
2006-19	8,03%	0,558	0,063	0,034	0,026
2006-20	8,52%	0,552	0,055	0,023	0,019
2006-21	8,05%	0,555	0,043	0,027	0,022
2006-22	7,76%	0,539	0,039	0,045	0,034



Obrázek 4: Stabilita sledovaných indexů v čase. Zdroj: vlastní konstrukce.

4. Závěr

Ačkoli vyvinutý prediktivní model může dosahovat skvělých charakteristik, jeho skutečná kvalita se prokáže až časem, tj. po jeho nasazení do praxe. Je tedy nezbytné výkonnost modelu pravidelně monitorovat. Vhodnými charakteristikami pro tyto účely jsou například Giniho index, K-S index, χ^2 statistika, PSI a PSI_{DR}. Zatímco první dva jmenované popisují především prediktivní sílu modelu, tj. např. schopnost rozlišit od sebe dobré a špatné klienty, další dva slouží k popisu a statistickému testování posunu distribuce vstupní populace skrze výstup prediktivního modelu, tj. např. skóre. Index PSI_{DR} reflektuje jak změnu vstupní populace, tak změnu absolutní hladiny default rate. Všechny uvedené charakteristiky je taktéž možné použít pro analýzu/monitoring jednotlivých vstupů modelu.

5. Literatura

[1]BRUNKE, O., MOUD, K.K., 2006, Credit Model Performance Monitoring System. [online] [cit. 2009-01-15] URL: <http://www.csrsa.org/files/conferences/2006/CreditModelPerformanceMonitoringSystem.ppt>.

[2] KARAKOULAS, G. 2004. Empirical Validation of Retail Credit-Scoring Models. The RMA Journal. ISSN 1531-0558

[3] LI, S., KHARIDHI, S., KRAMER, M., 2008, Using PSI to Monitor Predictive Model Stability in the Database Marketing Industry. SESUG Proceedings [online]. [cit. 2009-01-15] URL: <http://analytics.ncsu.edu/sesug/2008/BI-007.pdf>.

[4] XU, K., 2003. How has the literature on Gini's index evolved in past 80 years? [online]. [cit. 2009-01-15] URL: <http://economics.dal.ca/RePEc/dal/wparch/howgini.pdf>.

Adresa autora:

Martin Řezáč, Mgr. Ph.D.
Ústav statistiky a operačního výzkumu PEF MZLU
Zemědělská 1
613 00 Brno
rezac@mendelu.cz

Štatistika v školskej matematike
Statistics in School Mathematics

Ondrej Šedivý

Abstract: The article deals with statistics education at primary and secondary schools. It provides a review of students' knowledge of statistics and their readiness for university study.

Key words: Statistics education, reform of education in primary and secondary school

Kľúčové slová: Vyučovanie štatistiky, školská reforma na základných a stredných školách

1. Úvod

V každodennom živote sa stretávame s použitím matematiky. Matematiku používame takmer všade. Niekde používame jednoduché pojmy, vzorce, jednoduché výpočty, inde používame zložité vzťahy vyjadrené matematickým jazykom, alebo zložité výpočtové práce, matematiku používame pri príprave podkladových materiálov pre rozhodovanie, inde zasa používame matematiku pri navrhovaní nových výrobkov a rôznych technológií. Takmer v každom zložitejšom rozhodovaní sa opierame o rôzne matematické a matematicko – štatistické metódy. Dnes si nevieme predstaviť prácu riadiaceho pracovníka, aby sa pri rozhodovaní neopieral o rôzne analýzy spracované matematickými a informačnými metódami, pritom sa veľa razy opierajú o štatistiku.

Teda matematika poskytuje určité služby spoločenskej praxi a súčasne spoločenská prax stavala a stavia pred matematiku najprv priamo a neskôr sprostredkovane stále nové úlohy a problémy, čím ju spája so životom. Ďalej spoločenská prax podnecuje jej rozvoj a orientuje ju v tom či onom smere. Treba však zdôrazniť, že spoločenská prax preveruje správnosť záverov a tvrdení matematiky.

2. Vyučovanie matematiky na základných školách

Na vyššie uvedené skutočnosti treba žiakov pripraviť tak, aby boli schopní v živote používať matematiku vo svojej práci.

Dôležitou súčasťou školskej matematiky je oblasť kombinatoriky, pravdepodobnosti a štatistiky. Uvedené súčasti tvoria podľa Štátneho programu vzdelávania jeden okruh.

Na prvom stupni základnej školy tento tematický okruh sa objavuje len v podobe úloh. Žiaci takéto úlohy na 1. stupni ZŠ riešia manipulatívnou činnosťou s konkrétnymi objektami, pričom vytvárajú rôzne skupiny predmetov podľa určitých pravidiel (usporiadúvajú rôzne zoskupenia). Pozorujú frekvenciu výskytu určitých javov, udalostí a zaznamenávajú.

Na 2. stupni ZŠ sa vyučovanie pravdepodobnosti a štatistiky sústreďuje na pravdepodobnostné hry a pokusy, plánovitý zber údajov a ich usporiadanie, tvorbu stĺpcového diagramu, interpretáciu jednoduchých grafov a ich rozbor na konkrétnych úlohách, výpočet aritmetického priemeru viacerých čísel, kruhový diagram, spôsob vyhľadávania a systematického vypisovania možností, početnosť a relatívna početnosť javu, výpočet relatívnej početnosti, využívanie kombinatorických poznatkov žiakov pri výpočte relatívnej početnosti javov. V 9. ročníku ZŠ sa poznatky z tejto oblasti uzatvárajú najmä týmto: zber údajov a ich systemizácia, zobrazenie skupín údajov a tvorba grafov a diagramov, spôsoby vyhľadávania, systematické vypisovanie možností, objavovanie a opis systému, algebraizácia systému alebo počtu možností, pravdepodobnostné pokusy a ich štatistické spracovanie.

3. Vyučovanie štatistiky na gymnáziách

Výučba štatistiky na gymnáziách nadväzuje na poznatky z ekonomiky a pravdepodobnosti a sústreďuje sa do 3. ročníka. Obsahom tohto celku je: Štatistika ako súbor metód robiť rozumné rozhodnutia v prípadoch „neistoty“, východisko pre plánovanie a organizáciu. Štatistický súbor, rozsah súboru. Štatistický znak a jeho hodnota. Početnosť a relatívna početnosť pre jednotlivé hodnoty (intervaly hodnôt) štatistického znaku – frekvenčné tabuľky. Grafické spracovanie dát (histogram, , kruhový diagram, čiarové grafy lomené a hladké). Použitie vhodného softvéru (napr. EXCEL) pri grafickom spracovaní dát. Porovnanie štatistického znaku pre rôzne výberové súbory, formalizácia hypotéz a ich intuitívne hodnotenie. Čo vypovedajú o súbore stredná hodnota, modus, medián a rozptyl. Použitie kalkulačky a počítačového softvéru (napr. EXCEL) pri základných štatistických výpočtoch. Vážený priemer. Príklady situácií, v ktorých nie je vhodné použitie aritmetického priemeru (napr. priemerná rýchlosť). Normálne rozdelenie (situácie, v ktorých je vhodné, resp. nevhodné jeho použitie). Percentily. Príklady iných rozdelení početnosti (pravdepodobnosti). Vo 4. ročníku v predmete štatistika sa orientuje na prieskum verejnej mienky a štatistické výskumy. Výberový súbor, kedy možno výsledky získané z výberového súboru považovať za platné pre celý súbor. Možné chyby pri interpretácii výsledkov.

4. Záver

Príspevok musí obsahovať zoznam použitej literatúry. Vzor bibliografického odkazu na knižnú publikáciu a aj článok sú uvedené v tejto šablóne. Na záver prosíme uveďte meno autora (autorov) a adresy (vrátane e-adresy) podľa uvedeného vzoru. Prosíme autorov o dodržiavanie pokynov na úpravu príspevkov. V prípade nedodržania pokynov na úpravu si vydavateľ vyhradzuje právo neuverejniť takýto príspevok.

5. Literatúra

[1] HEJNÝ, M. a kol.: Základy teórie vyučovania matematiky 2. Bratislava, SPN, 1989. ISBN 80 - 08 - 00014 – 7

[2] ŠEDIVÝ, O. a kol.: Matematika pre 9. ročník základných škôl – 2. časť, Bratislava, SPN, 2005. ISBN 80 – 10 – 00857 – 5

[3] Štátny vzdelávací program – matematika, ŠPÚ, 2008

Adresa autora:

Prof. RNDr. Ondrej Šedivý, CSc.
Katedra matematiky FPV UKF
Trieda A. Hlinku 1
949 74 Nitra
osedivy@ukf.sk

Využitie fuzzy metód pri určení rizikovej skupiny drogovej závislosti
Utilization of Fuzzy Methods for Determine Endangered Groups of Drug Dependence

Beáta Stehlíková, Jaroslav Ivor, Anna Tirpáková, Dagmar Markechová

Abstract: The contribution describes drug use in general population. To observe "vulnerable" groups of population on drug, we used statistical methods.

Key words: Drug prevalence, drug use, fuzzy methods

Kľúčové slová: Rozšírenie drog, užívanie drog, fuzzy metódy

1. Úvod

Štatistický úrad SR v [6] uvádza, že droga je látka, ktorá je po vstupe do živého organizmu schopná pozmeniť jednu alebo viac jeho funkcií, pôsobiť priamo alebo nepriamo na centrálny nervový systém. Môže mať priznané postavenie lieku. Miera rizika vzniku závislosti slúži ako hlavné kritérium pre delenie drog na tzv. mäkké a tvrdé alebo ľahké a ťažké. Za ľahké, mäkké drogy sa považujú tabakové výrobky všetkých druhov, káva, produkty z konope - kanabis a alkohol. Kanabis je najrozšírenejšou drogou, buď v rastlinnej forme ako marihuana (tráva), alebo ako lisovaná živica (hašiš), či ako vysoko aktívny konopný olej. Riziko závislosti od kanabisu je menšie ako napríklad od heroínu. Medzi tvrdé, ťažké drogy patria návykové látky patriace do skupiny opiátov zaradených do širšej skupiny narkotík. Ide predovšetkým o heroín, morfín a kodeín.

Užívanie drog je jedným z piatich kľúčových indikátorov Európskeho monitorovacieho centra pre drogy a drogovú závislosť (EMCDDA), ktorý sa používa na deskripciu stavu v užívaní legálnych a nelegálnych látok v danej krajine. Podľa EMCDDA [2] za problémové užívanie drog je tu považované injekčné alebo dlhodobé resp. pravidelné užívanie opiátov, kokaínu a/alebo amfetamínov vo vekovej skupine 15 – 64 rokov v danom roku.

Rizikom opiátov je predávkovanie a smrť- pri zvyšujúcich sa dávkach, či neznalosti kvality drogy, fyzická a psychická závislosť s ťažkými abstinenčnými príznakmi, ktoré nastupujú krátko po odznení účinku dávky. Heroín vystupuje spomedzi ostatných drog ako najčastejšia príčina úmrtí v súvislosti s drogami, uvádza sa tiež v [8].

S užívaním drog súvisí aj trestná činnosť. Prezídium policajného zboru (PPZ) zaviedlo povinné vykazovanie druhov drog pri spáchaní trestných činov podľa § 171 – 174 Trestného zákona od 1. júna 2006. V roku 2007 PPZ po druhýkrát viedlo štatistické sledovanie počtu stíhaných páchateľov a trestných činov podľa druhu drogy. Evidoval celkom 2 160 trestných činov a 1 861 stíhaných páchateľov, uvádza sa vo Výročnej správe o stave drogovej problematiky na Slovensku za rok 2007 [7].

Kto je najviac ohrozený? Sú to takzvané **z**raniteľné skupiny z hľadiska užívania drog- ide o konkrétne skupiny spomedzi širšej populácie, ktoré môžu byť náchylnejšie k celému radu problémov, počnúc zlým zdravotným stavom až po horšie študijné výsledky. Sociálno – ekonomická situácia občanov tiež do určitej miery ovplyvňuje užívanie nelegálnych drog. Tak ako uvádza Výročná správa, aj ďalšie štatistiky za rizikovú skupinu jednoznačne považujú skupinu populácie do 25 rokov a so zvyšujúcim sa vekom miera rizika užívania

drog klesá. Z toho dôvodu sú prevencie zamerané predovšetkým na mládež. Je pravda, že miera rizika užívania drog s rastúcim vekom klesá?

2. Materiál a metódy

Pojem Fuzzy Matrix Model zaviedol Kandasamy et al. [4], a to nasledujúcim spôsobom. Východiskom sú vstupné údaje zapísané do matice TDM (Time Dependent Matrix). Prvky v každom riadku sa následne znormujú počtom rokov, čím dostaneme maticu s prvkami a_{ij} (i = 1, 2, …, r; j = 1, 2, …, c), kde r je počet vekových skupín a c je počet druhov drog. Takto upravená matica sa konvertuje do matice ATD (Average Time Dependent Data Matrix) s prvkami e_{ij} (i = 1, 2, …, r; j = 1, 2, …, c), kde r je počet vekových skupín a c je počet druhov drog. Prvky e_{ij} (i = 1, 2, …, r; j = 1, 2, …, c) vypočítame pomocou nasledujúcich vzťahov

$$e_{ij} = \begin{cases} -1, & pre\ a_{ij} \le \mu_j - \alpha\sigma_j, \\ 1, & pre\ a_{ij} \ge \mu_j + \alpha\sigma_j, \\ 0, & inde, \end{cases}$$

kde μ_j (j = 1, 2, ..., c) je priemer za j-ty stĺpec a σ_j (j = 1, 2, ..., c) je smerodajná odchýlka údajov j-teho stĺpca, α je parameter z intervalu $\langle 0, 1 \rangle$. Sčítaním prvkov matice v každom riadku dostaneme stĺpcové vektory **v** , tzv. maticu CETD (Combined Effective Time Dependent Data Matrix). Východiskom výpočtov bola tabuľka prevalencie drog za rok 2006 podľa pohlavia a veku [9].

Tabuľka 1: Všeobecné rozšírenie drog za rok 2006 podľa veku v % (muži)

Vekové kategórie	Drogy				
	kanabis	heroin	kokain	amfetamín	extáza
15-24	39	3,3	4	5,3	11,8
25-34	39	1,5	3	1,5	9,6
35-44	19,1	0,9	1,8	0,9	5,4
45-54	7,1	0	0	0,8	0,8
55-64	4,1	0	0	0	0

Zdroj. [9]

Tabuľka 2: Všeobecné rozšírenie drog za rok 2006 podľa veku v %(ženy)

Vekové kategórie	Drogy				
	kanabis	heroin	kokain	amfetamín	extáza
15-24	22,3	1,5	0,8	1,5	7,7
25-34	12	0,8	0	0,8	3,8
35-44	6	0,7	1,4	0	2
45-54	3,9	0	0,6	0	0
55-64	1,9	0	0	0	

Zdroj. [9]

Tabuľka 3: Všeobecné rozšírenie drog za rok 2006 podľa veku celkom

Vekové kategórie	Drogy				
	kanabis	heroin	kokain	amfetamín	extáza
15-24	31,3	2,5	2,5	3,6	9,9
25-34	25,7	1,1	1,5	1,1	6,7
35-44	11,6	0,8	1,6	0,4	3,5
45-54	5,3	0	0,4	0,4	0,4
55-64	3,0	0	0	0	0

Zdroj. [9]

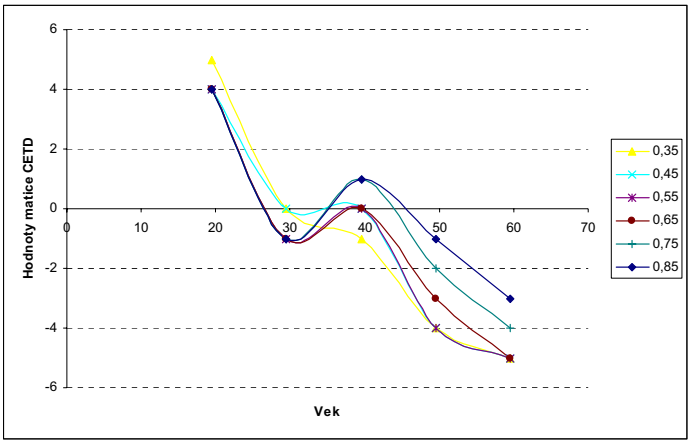
3. Výsledky a diskusia

Po výpočtoch pre zvolené hodnoty parametra α sme dostali nasledovné hodnoty matíc CETD (Tabuľka 4,5,6), ktoré sme pre ilustráciu znázornili aj graficky (Graf 1,2,3)

Tabuľka 4: Hodnoty matice CETD (muži)

Vekové kategórie	α = 0,35	α = 0,45	α = 0,55	α = 0,65	α = 0,75	α = 0,85
15-24	5	5	5	5	5	5
25-34	3	3	3	3	2	1
35-44	-1	0	0	0	0	0
45-54	-5	-4	-4	-4	-4	-3
55-64	-5	-5	-5	-5	-5	-3

Zdroj: vlastný výpočet



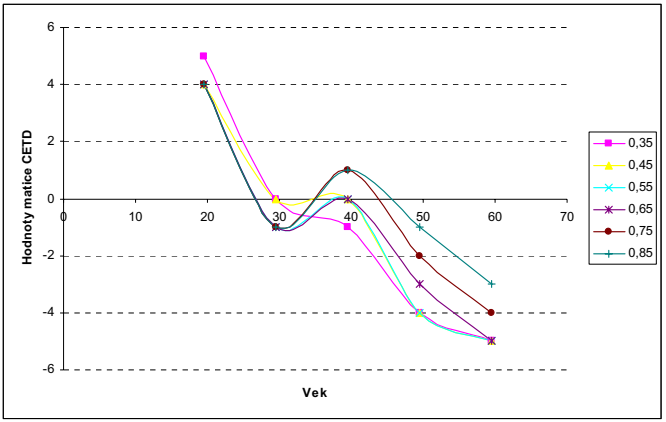
Zdroj: vlastné zobrazenie

Obrázok 1: Rizikové skupiny pre drogovú závislosť(muži)

Tabuľka 5: Hodnoty matice CETD (ženy)

Vekové kategórie	α = 0,35	α = 0,45	α = 0,55	α = 0,65	α = 0,75	α = 0,85
15-24	5	4	4	4	4	4
25-34	0	0	-1	-1	-1	-1
35-44	-1	0	0	0	1	1
45-54	-4	-4	-4	-3	-2	-1
55-64	-5	-5	-5	-5	-4	-3

Zdroj: vlastný výpočet



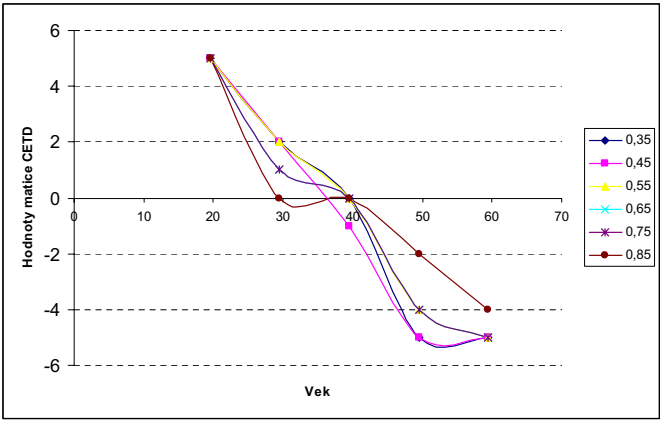
Zdroj: vlastné zobrazenie

Obrázok 2: Rizikové skupiny pre drogovú závislosť(ženy)

Tabuľka 6: Hodnoty matice CETD (spolu)

Vekové kategórie	α = 0,35	α = 0,45	α = 0,55	α = 0,65	α = 0,75	α = 0,85
15-24	5	5	5	5	5	5
25-34	2	2	2	1	1	0
35-44	0	-1	0	0	0	0
45-54	-5	-5	-4	-4	-4	-2
55-64	-5	-5	-5	-5	-5	-4

Zdroj: vlastný výpočet



Zdroj: vlastné zobrazenie

Obrázok 3: Rizikové skupiny pre drogovú závislosť(spolu)

Z obrázkov vidíme, že prezentovaný postup je stabilný vzhľadom na voľbu parametra α. Na základe výsledkov, ktoré sme dostali použitím uvedenej fuzzy metódy, môžeme konštatovať, že rizikové skupiny z hľadiska užívania drog sú dve. Metóda potvrdila, že skutočne rizikovou skupinou je mládež, ale tiež ukázala, že druhou rizikovou skupinou je populácia vo veku od 35 do 44 rokov. Na grafoch vidíme, že krivka monotónne neklesá, t.j. nie je pravda, že miera rizika užívania drog s rastúcim vekom klesá. Táto skutočnosť platí tak pre ženy ako aj mužov.

4. Záver

Uvedenou metódou sa podarilo dokázať, že okrem rizikovej skupiny mladých ľudí do 25 rokov existuje aj ďalšia riziková skupina populácie, a to sú ľudia vo veku od 35 do 44 rokov. Možno by bolo vhodné zamerať v budúcnosti prieskum kompetentných orgánov aj na túto vekovú kategóriu populácie, aby sa zistili príčiny nárastu a dôvody užívania drog a tak sa mohla zvoliť vhodná prevencia aj pre túto vekovú kategóriu.

5. Literatúra

[1]ANNUAL REPORT (2008): the state of the drugs problem in Europe. EMCDDA, Lisbon, November 2008, Publication type: Annual report
http://www.emcdda.europa.eu/publications/country-overviews/ge?LayoutFormat=print

[2]EMCDDA: Key epidemiological indicator: Prevalence of problem drug use. EMCDDA recommended draft technical tools and guidelines. Lisabon, EMCDDA, 2004, http://www.emcdda.europa.eu, http://www.emcdda.europa.eu/html.cfm/index190EN.html

[3]EURÓPSKE MONITOROVACIE CENTRUM PRE DROGY A DROGOVÚZÁVISLOSŤ. Výročná správa 2008: stav drogovej problematiky v Európe. Luxemburg: Úrad pre vydávanie úradných publikácií Európskych spoločenstiev (2008) 98 s. ISBN 978-92-9168-338-3

[4]KANDASAMY, W.B.V. – SMARANDACHE, F. - IIANTHENRAL, K.: Elementary Fuzzy Matrix Theory and Fuzzy Models for Social Scientists. Los Angeles 2007, ISBN(10): 1-59973-005-7, ISBN(13): 978-1-59973-005-9

[5]NÁRODNÉ MONITOROVACIE CENTRUM PRE DROGY: 2008 Národná správa (2007 údaje) pre EMCDDA. Bratislava 2008, ŠEVT a. s., ISBN 978-80-8106-007-6

[6]ŠTATISTICKÝ ÚRAD SLOVENSKEJ REPUBLIKY: Výročná správa o stave drogovej problematiky na Slovensku za rok 2006, http://portal.statistics.sk/showdoc.do?docid

[7]ŠTATISTICKÝ ÚRAD SLOVENSKEJ REPUBLIKY: Výročná správa o stave drogovej problematiky na Slovensku za rok 2007, http://portal.statistics.sk/showdoc.do?docid

[8] http://portal.statistics.sk/showdoc.do?docid

[9] http://portal.statistics.sk/showdoc.do?docid=5431

Adresy autorov:

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

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

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

Dagmar Markechová, Doc. RNDr., CSc.
Katedra Matematiky UKF
Tr. A. Hlinku 1
949 01 Nitra1
dmarkechova@ukf.sk

Hodnotenie efektívnosti inovačnej výkonnosti štátov EÚ
The Evaluation of Effectiveness of Innovative Efficiency of EU States

Anna Tirpáková, Beáta Stehlíková, Dagmar Markechová, Jozef Stieranka

Abstract: The article evaluates the innovation process in countries of European Union by using mathematical - statistical methods. The innovation index used by the European Commission is a tool for evaluating the innovation process in countries of European Union

Key words: Innovation process, Economic growth, Mathematical - statistical methods
Kľúčové slová: Inovačný proces, ekonomický rast, matematicko-štatistické metódy

1. Úvod

Je známe, že ekonomický rast je možné dosiahnuť kvantitatívne navršovaním kapitálových investícií. Efektívnejším spôsobom sú však investície do technologického pokroku, ktoré sa prejavia zvýšením produktivity práce. Pre ekonomiku Slovenska sú pre proces inovácií dôležité jednak poznatky prinesené zahraničnými investormi, ale aj vlastný inovačný proces. Európska komisia na hodnotenie jednotlivých kľúčových determinantov ovplyvňujúcich inovačný proces v krajinách EÚ používa index inovácie, ktorý obsahuje päť kategórií. Tieto kategórie pozostávajú z 26 indikátorov inovačného procesu. Na základe rozdelenia indikátorov je možné identifikovať silné a slabé stránky inovačného procesu jednotlivých štátov EÚ. Cieľom je dosiahnuť rovnomerný rozvoj zložiek inovačného indexu, ktorý sa následne prejaví zvýšením inovačnej výkonnosti štátov.

Predložený príspevok je zameraný na hodnotenie inovačného potenciálu, a to z toho dôvodu, že rok 2009 bol vyhlásený rokom tvorivosti a inovácie s oficiálnym sloganom „Myslíme. Tvoríme. Inovujeme.". Európsky rok tvorivosti a inovácie je zameraný na zvýšenie povedomia o význame tvorivosti a inovácie ako kľúčových spôsobilostí pre osobný, spoločenský a hospodársky rozvoj. Cieľom je podporovať tvorivé a inovačné prístupy v rozličných odvetviach ľudskej činnosti a prispievať k lepšej príprave Európskej únie na budúce výzvy v globalizovanom svete, uvádza sa v [9]. Príspevok obsahuje hodnotenie inovačného procesu v krajinách EÚ pomocou matematicko-štatistických metód.

2. Materiál

EIS (European Innovation Scoreboard) je nástrojom Európskej komisie pre hodnotenie a porovnávanie inovačnej výkonnosti štátov Európskej únie. EIS 2005 obsahuje sumárny inovačný index SII (Summary Innovation Index) a analýzu inovačného trendu. Indexu inovácie SII EIS je venovaná značná politická pozornosť. Týchto 26 indikátorov inovácie je rozdelených do piatich kategórií, ktoré pokrývajú rozdielne kľúčové dimenzie inovačnej činnosti. V kategórii *Hnacie sily inovácie* je päť indikátorov, ktoré sú mierou štrukturálnych podmienok potrebných pre schopnosť inovácie. Kategória *Tvorba poznatkov* obsahuje tiež päť indikátorov a je mierou investícií do výskumu a vývoja. Úsilie k inovácii na firemnej úrovni je kvantifikované šiestimi indikátormi tvoriacimi kategóriu *Inovácia a súkromné podnikanie*. Päť indikátorov kategórie *Aplikácie* merajú výkonnosť pracovných a obchodných aktivít a ich pridanú hodnotu do inovačných sektorov. Kategória *Duševné vlastníctvo* obsahuje päť indikátorov a je mierou dosiahnutých výsledkov v termínoch úspešných know-how. Týchto päť kategórií pokrýva jednotlivé aspekty inovačného procesu. Indikátory zaradené do prvých troch kategórií sa týkajú vstupu a zvyšné dve charakterizujú výstup

inovačného procesu. Členenie indikátorov do uvedených kategórií poskytuje rýchlu identifikáciu silných a slabých stránok jednotlivých štátov v inovačnom procese. Nasledujúca tabuľka obsahuje prehľad kategórií inovačného indexu SII.

Tabuľka 1: Kategórie sumárneho inovačného indexu SII

Kategória	Charakteristika
1. hnacie sily inovácie (5 indikátorov)	je mierou štrukturálnych podmienok potrebných pre schopnosť inovácie
2. tvorba poznatkov (5 indikátorov)	je mierou investícií do výskumu a vývoja
3. inovácia a súkromné podnikanie (6 indikátorov)	meria úsilie k inovácii na firemnej úrovni
4. aplikácie (5 indikátorov)	merajú výkonnosť pracovných a obchodných aktivít a ich pridanú hodnotu do inovačných sektorov
5. duševné vlastníctvo (5 indikátorov)	je mierou dosiahnutých výsledkov v termínoch úspešných know-how

Zdroj: [16]

3. Metódy

Pre hodnotenie variability sa používa smerodajná odchýlka, ktorá je však závislá od mernej jednotky. Variačný koeficient tento nedostatok odstraňuje, ale nadobúda hodnoty z oboru reálnych čísiel. Na odstránenie uvedeného nedostatku možno použiť Giniho koeficient [13]

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|}{2n^2\bar{x}},$$

kde x_i je hodnota meraného znaku, n je počet hodnotených subjektov.

Jednou z možnosti ako vyhodnotiť efektívnosť jednotlivých ekonomických subjektov v rámci určitej skupiny je použitie analýzy DEA (Data Envelopment Analysis) [7].

DEA analýzu je možné aplikovať aj na hodnotenie efektívnosti štátov Európskej únie z hľadiska investícií. Vstupom sú hodnoty prvých troch indikátorov - hnacie sily inovácie (1), tvorba poznatkov (2) a inovácia a súkromné podnikanie (3). Výstup predstavujú hodnoty indikátorov aplikácie (1) a duševné vlastníctvo (2).

Existuje viacero typov DEA modelov. Ich cieľom je odlíšiť efektívne subjekty od neefektívnych. Súčasne DEA umožňuje určiť mieru efektívnosti neefektívnych subjektov. DEA modely spoľahlivo určia efektívne subjekty ako útvary na efektívnej hranici. Nevýhodou niektorých DEA modelov je, že miery efektívnosti, ktoré poskytujú, nie sú vhodné na porovnávanie neefektívnych útvarov medzi sebou. Táto situácia nastáva najmä pri aplikácii DEA modelov na veľké súbory hodnotených subjektov. Preto má zmysel hľadať také miery, ktoré umožnia lepšie porovnávať neefektívne útvary medzi sebou. Subjekty podrobujúce sa DEA analýze označujeme DMU (Decision Making Units).

Predpokladáme, že subjekty sa zaoberajú rovnakou činnosťou, ktorú možno charakterizovať určitým počtom vstupov a určitým počtom výstupov. Vstupy sú také veličiny, ktoré sa pri danej činnosti spotrebovávajú. Pre vstupy sa vyžadujú čo najmenšie hodnoty. Výstupmi označujeme veličiny, ktoré sú produktom danej činnosti a majú byť pokiaľ možno čo najväčšie.

Budeme predpokladať, že sa hodnotí p subjektov z hľadiska činnosti, ktorá je charakterizovaná m vstupmi a n výstupmi. Nech $x_{oi} = (x_{o1}, x_{o2}, ..., x_{om})^T$ kde $i = 1,...,m$ je hodnota i-teho vstupu o-teho hodnoteného subjektu a $y_{ok} = (y_{o1}, y_{o2}, ..., y_{on})^T$ kde $k = 1,...,n$, je hodnota k-teho výstupu o-teho subjektu. Nech hodnoty vstupov a výstupov sú pre každý subjekt nezáporné, pričom každý subjekt má hodnotu aspoň jedného vstupu a aspoň jedného výstupu nenulovú.

Aditívny model s variabilnými výnosmi z rozsahu VRTS (variable return to scale) predpokladá riešiť pre pevne zvolené $o \in \{1, 2, ..., p\}$ nasledovnú úlohu lineárneho programovania v primárnom tvare:

$$\min_{\boldsymbol{\lambda}, \mathbf{s}^+, \mathbf{s}^-} -((\mathbf{w}^-)^T \mathbf{s}^- + (\mathbf{w}^+)^T \mathbf{s}^+),$$

za podmienok:

$$\sum_{j=1}^{p} \lambda_j \mathbf{x}_j + \mathbf{s}^- = \mathbf{x}_0, \ \sum_{j=1}^{p} \lambda_j \mathbf{y}_j - \mathbf{s}^+ = \mathbf{y}_0, \ \sum_{j=1}^{p} \lambda_j = 1, \ \lambda_j \geq 0, \ \ j = 1, 2, ..., p, \ \mathbf{s}^+ \geq 0, \mathbf{s}^- \geq 0,$$

kde $\mathbf{s}^-$, $\mathbf{s}^+$ sú m a n rozmerné vektory doplnkových vstupných a výstupných premenných. $\mathbf{w}^-$, $\mathbf{w}^+$ sú m a n rozmerné vektory váh doplnkových vstupných a výstupných premenných. Nech usporiadaná dvojica vektorov $(\mathbf{u}^*, \mathbf{v}^*)$ je optimálnym riešením, E_0^* je optimálna hodnota účelovej funkcie. Ak sa hodnota E_0^* rovná nule, tak hodnotený subjekt je efektívny, inak je neefektívny. Hodnota E_0^* je určitou mierou neefektívnosti. Jej nevýhodou je, že je zdola neohraničená. Vektory $\mathbf{w}^-$, $\mathbf{w}^+$ je možné nahradiť vektormi pozostávajúcimi zo samých jednotiek. Normované vektory váh sa volia ako recipročná hodnota smerodajnej odchýlky σ_i^- i-tého vstupu, resp. smerodajnej odchýlky σ_i^+ i-tého výstupu, t.j. $w_i^- = 1/\sigma_i^-$ $i = 1, 2, ..., m$, $w_i^+ = \sigma_i^+$ $i = 1, 2, ..., n$. Parciálne efektívnosti sú definované pomocou vzťahov

$$E_i^{vstup} = \frac{x_{oi} - s_i^-}{x_{oi}}, \ i = 1, 2, ..., m, \ E_i^{výstup} = \frac{y_{oi} - s_i^+}{y_{oi}} \ i = 1, 2, ..., n.$$

Agregovaná miera efektívnosti je

$$E = \sum_{i=1}^{m} a_i \ E_i^{vstup} + \sum_{i=1}^{n} b_i \ E_i^{výstup}, \text{ pričom } \sum_{i=1}^{m} a_i + \sum_{i=1}^{n} b_i = 1.$$

Váhy a_i (i = 1, 2, ..., m) a b_i (i = 1, 2, ..., n) je možné voliť viacerými spôsobmi. V prípade rovnakých váh $a_i = b_i = 1/(m + n)$. Iná možnosť je voľba pomocou váh w, t.j.

$$a_i = \frac{w_i^-}{\sum w_i^- + \sum w_i^+}, \text{ i} = 1, 2, ..., m, \ b_i = \frac{w_i^+}{\sum w_i^- + \sum w_i^+}, \text{ i} = 1, 2, ..., m,$$

Miera efektívnosti nadobúda hodnoty z intervalu $\langle 0, 1 \rangle$ a je invariantná vzhľadom na vynásobenie kladnou konštantou, ale nie je invariantná vzhľadom na pričítanie konštanty k i-tému vstupu, resp. výstupu [31].

4. Výsledky a diskusia

V nasledujúcich tabuľkách sú uvedené hodnoty kategórií sumárneho inovačného indexu (Tabuľka 2) a výsledky DEA analýzy (Tabuľka 3a Tabuľka 3b).

Tabuľka 2: Hodnoty kategórií sumárneho inovačného indexu SII

Územie	Hnacie sily inovácie	Tvorba poznatkov	Inovácia a súkromné podnikanie	Aplikácie	Duševné vlastníctvo	SII
EU25	0,4648	0,5339	0,4133	0,5190	0,3562	0,4574
EU15	0,4912	0,5956	0,4532	0,6044	0,4184	0,5125
Slovensko	0,2950	0,0757	0,1915	0,5539	0,0169	0,2266

Zdroj: [3]

Tabuľka 3a: Výsledky DEA analýzy

Štát	Účelová funkcia	Doplnkové premenné				
		s_1	s_2	s_3	$-s_1$	$-s_2$
Belgicko	-4,315	0,353	0,048	0,118	0,247	0,000
Česká republika	-3,356	0,169	0,116	0,123	0,153	0,056
Dánsko	-4,491	0,429	0,000	0,202	0,168	0,000
Estónsko	-4,966	0,241	0,000	0,410	0,203	0,003
Grécko	-4,341	0,112	0,000	0,091	0,479	0,067
Francúzsko	-4,217	0,386	0,130	0,069	0,184	0,000
Írsko	-4,920	0,385	0,228	0,103	0,187	0,000
Lotyšsko	-5,789	0,142	0,000	0,239	0,557	0,049
Litva	-6,230	0,329	0,000	0,176	0,512	0,077
Maďarsko	-2,596	0,046	0,207	0,000	0,201	0,024
Rakúsko	-0,910	0,006	0,027	0,000	0,121	0,000
Poľsko	-3,058	0,100	0,075	0,000	0,331	0,031
Slovinsko	-4,557	0,354	0,125	0,071	0,247	0,033
Fínsko	-4,627	0,455	0,192	0,131	0,081	0,000
Švédsko	-6,162	0,449	0,251	0,323	0,102	0,000
Spojené kráľovstvo	-4,505	0,486	0,079	0,150	0,108	0,000
Bulharsko	-3,103	0,129	0,015	0,000	0,338	0,069

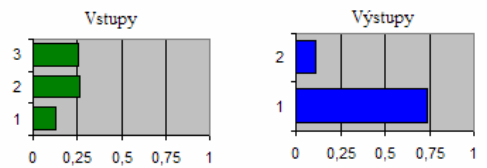
Zdroj: Vlastné výpočty

Na základe výsledkov možno konštatovať, že z hľadiska vstupov a výstupov inovácií je v Európskej únii efektívnych deväť štátov - Nemecko, Malta, Holandsko, Portugalsko, Španielsko, Taliansko, Luxembursko, Rumunsko a Slovensko. Výsledok DEA analýzy je v Tabuľke 3a a v Tabuľke 3b. Tabuľky obsahujú hodnoty doplnkových premenných a hodnoty λ. Napríklad hodnota λ pre Českú republiku je maximálna (dokonca jediná) korešpondujúca so štátom Malta. Na obrázkoch 1 a 2 sú znázornené rezervy vstupov a výstupov pre Maltu a Českú republiku.

Tabuľka 3b: Výsledky DEA analýzy

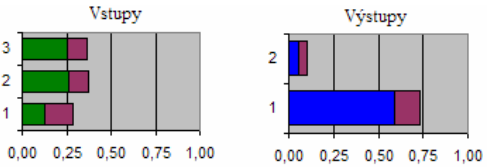
	Hodnoty lambda					
Štát	**Malta**	**Slovensko**	**Nemecko**	**Luxembursko**	**Holandsko**	**Taliansko**
Belgicko	0,588		0,412			
Česká republika	1,000					
Dánsko	0,204		0,728	0,067		
Estónsko	0,165	0,835				
Grécko	0,740	0,260				
Francúzsko	0,634		0,366			
Írsko	0,748		0,252			
Lotyšsko	0,524	0,476				
Litva	0,767	0,233				
Maďarsko	0,415	0,585				
Rakúsko			0,437		0,313	0,250
Poľsko	0,453	0,547				
Slovinsko	1,000					
Fínsko	0,167		0,833			
Švédsko	0,177		0,823			
Spojené kráľovstvo	0,635		0,365			
Bulharsko	0,540	0,460				

Zdroj: Vlastné výpočty



Zdroj: Vlastné výpočty

Obrázok 1: Vstupy a výstupy efektívneho štátu Malta



Zdroj: Vlastné výpočty

Obrázok 2: Vstupy a výstupy neefektívnej Českej republiky s vyznačenými rezervami

Tabuľka 4: Parciálne efektívnosti vstupov a výstupov

Štát	E_1^{vstup}	E_2^{vstup}	E_3^{vstup}	$E_1^{výstup}$	$E_2^{výstup}$
Belgicko	0,4038	0,8953	0,7592	0,6570	1,0000
Česká republika	0,4295	0,6949	0,6737	0,7937	0,5083
Dánsko	0,4433	1,0000	0,7066	0,7562	1,0000
Estónsko	0,5260	1,0000	0,3299	0,6529	0,9227
Grécko	0,6036	1,0000	0,7235	0,3089	0,2421
Francúzsko	0,3702	0,7521	0,8384	0,7453	1,0000
Írsko	0,3370	0,6079	0,7603	0,7431	1,0000
Lotyšsko	0,5933	1,0000	0,4837	0,1463	0,2757
Litva	0,3359	1,0000	0,5762	0,2668	0,1507
Maďarsko	0,8309	0,4264	1,0000	0,6822	0,5812
Rakúsko	0,9857	0,9542	1,0000	0,7722	1,0000
Poľsko	0,6870	0,6813	1,0000	0,4815	0,4891
Slovinsko	0,2640	0,6779	0,7824	0,6670	0,7055
Fínsko	0,4375	0,7447	0,7898	0,8834	1,0000
Švédsko	0,4386	0,6890	0,6028	0,8534	1,0000
Spojené kráľovstvo	0,3179	0,8322	0,7055	0,8508	1,0000
Bulharsko	0,6126	0,9237	1,0000	0,4842	0,0037

Zdroj: Vlastné výpočty

Tabuľka 5: Agregované miery efektívnosti pre jednotlivé štáty a ich poradia

	Agregovaná miera efektívnosti				
Štát	**E */**	**Poradie */**	**E /**	**Poradie **/**	**Rozdiel poradí**
Rakúsko	0,9424	1	0,9352	1	0
Dánsko	0,7812	2	0,7651	2	0
Fínsko	0,7711	3	0,7644	3	0
Belgicko	0,7431	4	0,7262	6	-2
Spojené kráľovstvo	0,7413	5	0,7302	5	0
Francúzsko	0,7412	6	0,7310	4	2
Švédsko	0,7167	7	0,7047	8	-1
Maďarsko	0,7041	8	0,7204	7	1
Írsko	0,6897	9	0,6790	9	0
Estónsko	0,6863	10	0,6605	11	-1
Poľsko	0,6678	11	0,6789	10	1
Česká republika	0,6200	12	0,6320	13	-1
Slovinsko	0,6194	13	0,6204	14	-1
Bulharsko	0,6048	14	0,6372	12	2
Grécko	0,5756	15	0,5821	15	0
Lotyšsko	0,4998	16	0,4927	16	0
Litva	0,4659	17	0,4715	17	0

*Poznámky: Váhy poradia tabulky4 - */ Rovnaké váhy 0,2; **/ váhy a, b pomocou váh w*
Zdroj: Vlastné výpočty

Tabuľka 6: Váhy agregovanej miery efektívnosti

Váhy	a_1	a_2	a_3	b_1	b_2
Rovnaké váhy	0,20000	0,200000	0,200000	0,200000	0,200000
Pomocou váh w	0,200103	0,188655	0,226503	0,233994	0,150744

Zdroj: Vlastné výpočty

Výsledkom analýzy inovačnej výkonnosti štátov sú parciálne efektívnosti vstupov a výstupov (Tabuľka 4) a agregované miery efektívnosti uvedené (Tabuľka 5). Agregované miery efektívnosti sa zmenou váh a_i (i = 1, 2, ..., *m*) a b_i (i = 1, 2, ..., *n*) podstatne nelíšia.

Na prvých troch miestach sú v prípade oboch typov váh Rakúsko, Dánsko a Fínsko (Tabuľka 5). Vyspelé krajiny EÚ 15 sú podľa agregovaných mier efektívnosti na začiatku rebríčka hodnotenia.

5. Záver

Použitá metóda DEA bola pôvodne vyvinutá na meranie efektívnosti neziskových organizácií ako sú školy, nemocnice, štátna a verejná správa. Aplikácia DEA by mohla byť prostriedkom objektívneho hodnotenia citlivých problémov, ktoré rezonujú aj v našej spoločnosti. V súčasnosti viac menej prevládajú politicky podfarbené stanoviská. Cieľom DEA (Data Envelopment Analysis) je z daného súboru organizačných jednotiek (t.j. výrobných aj nevýrobných subjektov) vybrať tie, ktoré sú efektívne. DEA okrem rozdelenia jednotiek na efektívne a neefektívne umožňuje pre neefektívne organizačné jednotky identifikovať zdroj neefektívnosti a určit tak spôsob, akým môže jednotka dosiahnúť hranicu efektívnosti prostredníctvom redukcie resp. navýšenia vstupov alebo výstupov.

Napríklad, v oblasti zdravotníctva je potrebná zvýšená prehľadnosť, spoľahlivosť a dostupnosť údajov o kvalite zdravotníckej starostlivosti. Indikátory kvality sú každoročne vypracovávané Ministerstvom zdravotníctva SR. Slovensko sa v rámci Európy radí na 22. miesto z 31 štátov v hodnotení systému zdravotníckej starostlivosti - čo sa týka jeho prístupnosti užívateľom. Tento poznatok bol zverejnený Európskym Zdravotníckym Uživateľským Indexom (EHCI) za rok 2008. Management nemocnice sleduje desiatky parametrov, které hodnotia efektivnosť procesov v nemocnici. Obvykle však jediným kritériom efektívnosti je iba zisk – bez exaktného kvantifikovania prepojenia výstupu-zisku na vstupy. Uvedené metódy hodnotia stav zdravotníckej starostlivosti. Aplikácia DEA metódy by však odokryla akým spôsobom by sa dala dosiahnuť hranica efektívnosti - v prípade, že ju nedosahujeme. Známe sú aplikácie hodnotenia efektívnosti nemocníc pomocou DEA - [23], [25], [6]. Kvalitu primárnej zdravotnej starostlivosti pomocou DEA hodnotí [27].

Zabezpečenie kvality je tiež jednou z ústredných tém prebiehajúcej reformy vysokého školstva na európskej úrovni v rámci bolonského procesu. Aj efektívnosť pedagogickej práce sa dá merať [2]. Tiež finančné rozpočty škôl by mohli byť objektivizované [21], resp. aj alokácia zdrojov ako taká [24]. Efektívnosť austrálskych univerzít pomocou DEA je hodnotená v [1], v Spojenom kráľovstve v [5].

Jedným z hlavných cieľov EÚ je poskytovať občanom vysokú úroveň ochrany v oblasti slobody, bezpečnosti a spravodlivosti. V celej Európe je však možné pozorovať tendenciu uprednostniť rozpočtové obmedzenia pri rozhodnutiach o polícii, konštatuje sa vo Akčnom pláne EuroCOP−u na roky 2008 – 2011. Prípadé zvýšenie rozpočtu je obvykle terčom útoku určitých politikov. Využitie DEA by umožnilo na Slovensku, podobne ako tomu bolo v prípade Anglicka a Welsu urobiť kvalifikované rozhodnutia o optimálnej veľkosti a štruktúre policajného zboru [15]. Metóda DEA by umožnila exaktnou metódou odkryť rezervy, resp. kvantifikovať nedostatok zdrojov. Metóda DEA bola aplikovaná na evaluáciu efektívnosti

zvládnutia kriminality v Anglicku aj v [33]. Metóda DEA bola použitá aj v prípade hodnotenia efektívnosti práce polície v Španielsku [12], na Taivane [28] a v Indii [4]. V [4] je prezentovaný tiež spôsob komparácie efektívnosti rozličných zložiek systému trestného súdnictva pomocou analýzy DEA.

Využívanie výsledkov výskumu a vývoja v praxi je jednou z najproblematickejších oblastí Lisabonskej stratégie. Kvalitné ľudské zdroje sú základným predpokladom pre vykonávanie výskumu a vývoja na medzinárodne porovnateľnej úrovni. Zásoba ľudských zdrojov sa najlepšie odráža v počte nových absolventov vysokých škôl, obzvlášť v počte absolventov prírodných a technických vied a ich podiele na celkovom počte absolventov. Ľudské zdroje, ktoré sú vzdelané a zamestnané v oblasti vedy a techniky, sú základom pre znalostnú ekonomiku. Priamo prispievajú k expanzii výskumných a vývojových činností a rozvoju technologických inovácií.

Vývoj a výskum ako aj technologické inovácie sú nenahraditeľným a najväčším zdrojom vysoko kvalitných poznatkov. Tvoria nosný pilier rozvoja spoločnosti ako aj zvyšovania životnej úrovne občanov jednotlivých štátov.

6. Literatúra

[1] ABBOTTA, M. – DOUCOULIAGOS, C.: The efficiency of Australian universities: a data envelopment analysis. Economics of Education Review, Volume 22, Issue 1, February 2003, Pages 89-97

[2] ACS, L. Vyhodnocovanie efektívnosti pedagogickej práce na FMFI UK. Diplomová práca. Bratislava: FMFI UK, 2005

[3] ARUNDEL, A. – HOLLANDERS, H.: EXIS: An Exploratory Approach to inovation Scoreboards. http://www.trendchart.org/scoreboards2004/scoreboard_papers.cfm

[4] ARVIND VERMA - SRINAGESH GAVIRNENI: Measuring police efficiency in India: an application of Data Envelopment Analysis. Policing: An International Journal of Police Strategies & Management, 2006 , Volume: 29, Issue: 1, Page: 125 – 145

[5] ATHANASSOPOULOS, ANTREAS D. - SHALE, ESTELLE: Assessing the Comparative Efficiency of Higher Education Institutions in the UK by Means of Data Envelopment Analysis

[6] BANNICK R.R - OZCAN Y.A.: Efficiency analysis of federally funded hospitals: comparison of DoD and VA hospitals using data envelopment analysis: Health Serv Manage Res. 1995 May;8(2):73-85

[7] BASLEY, J.E.: Allocating fixed costs and resources via data envelopment analysis, European Journal of Operational Research, vol. 147, 2003, pp198-216

[8] BRAUNERHEILM, P. – JOHANSON, D.: Determinants of spatial concentration. Journal of Industry Studies, 2003

[9] Creativity and Innovation European Year 2009. Dostupné na internete: http://www.create2009.europa.eu/index_en.html

[10] ČUTKOVÁ, Z. – DONOVAL, M.: Odvetvová.: špecializácia v SR. BIATEC, 12, č.8, 2004, str. 6-9

[11] DENISON, E.: Trends in American Economic Growth, 1929-1982. The Brookings Institution, 1985

[12] DIEZ-TICIO, A. - MANCEBON, M. J.: The Efficiency of the Spanish Police Service: An Application of the Multiactivity DEA Model. Applied Economics February/2002, Volume: 34 , Issue: 3 , Pages: 351-62

[13] DIXON, P.M. –WEINER, J. – MITCHELL-OLD, T. – WOODLEY, R.: Bootstrapping the **Gini** coefficient of inequality. Ecology 68, 1987, p. 1548–1551

[14] DORNBUSCH, R. – FISCHER, S: Makroekonomie. Praha: SPN a Nadace Economics, 1994 ISBN 80-04-25 556-6

[15] DRAKE, L. - SIMPER, R.: Productivity estimation and the size-efficiency relationship in English and Welsh police forces: An application of data envelopment analysis and multiple discriminant analysis. International Review of Law and Economics, Volume 20, Issue 1, March 2000, Pages 53-73

[16] European Trend Chart on Innovation. http://trendchart.cordis.lu (30.08.2008)

[17] FABRICANT, S.: Economic Progress and Economic Change', 34th Report of the National Bureau of Economic Research, New York: NBER, 1954

[18] FANDEL, P. - BALÁŽOVÁ, E. :Hodnotenie efektívnosti poskytovania verejných služieb vo vybraných obciach SR aplikáciou DEA a FDH analýzy**.** Acta regionalia et environmentalica (online), roč. 5, 2008, č. 1, s. 6 – 13

[19] GABRIELOVÁ, H.: Pozícia Slovenska v porovnaní so štátmi EÚ v technologickej a poznatkovej náročnosti. Ústav slovenskej a svetovej ekonomiky SAV, Bratislava, 2005, 16. s.

[20] HAALAND, J. – KIND, H. – MIDELFART-KNARVIK, K. – TORSTENSON, J.: What determines the economic geography of Europe? Discusion paper 19/98 Bergen: EC, 1998

[21] CHALOS, P.: An Examination of Budgetary Inefficiency in Education Using Data Envelopment Analysis. Financial Accountability & Management, Volume 13 Issue 1, Pages 55 - 69

[22] Innovation Scoreboard 2005 http://www.cordis.lu/trendchart (30.08.2008)

[23] JACOBS, R.: Alternative Methods to Examine Hospital Efficiency: Data Envelopment Analysis and Stochastic Frontier Analysis, s. 103-115
Health Care Management Science, Volume 4, Number 2 / June, 2001

[24] LOZANO, S. - VILLA, G.: Centralized Resource Allocation Using Data Envelopment Analysis. s. 143-161, Journal of Productivity Analysis, Volume 22, Numbers 1-2 / July, 2004

[25] MAGNUSSEN, J.: Efficiency measurement and the operationalization of hospital production. Health Serv Res. 1996 April; 31(1): 21–37

[26] PUGA, D.: The rise and fall of regional inequalities. European economic review. 43,1999, pp.303-335

[27] SALINAS-JIMÉNEZ. J. – SMITH, P.: Data envelopment analysis applied to quality in primary health care. Annals of Operations Research, Volume 67, Number 1 / December, 1996, str. 141-161

[28] SHINN SUN: Measuring the relative efficiency of police precincts using data envelopment analysis . Socio-Economic Planning Sciences, Volume 36, Issue 1, March 2002, Pages 51-71

[29] SOLOW, R.: Technical Change and the Aggregate Production function. Review of Economics and Statistics, 1957

[30] STEHLÍKOVÁ, B.: Európska únia v zrkadle čísel. Nitra: Environment, 2008, 321 s. ISBN 978-80-969120-9-4

[31] ŠEVČIVIČ, M. – HALICKÁ, M. – BRUNOVSKÝ, P.: DEA analysis for a large structured bank branch network, Central Euro. J. of Operational Res. 9 (2001), 329--342

[32] ŠIKULOVÁ, I..: Štrukturálna konvergencia Slovenskej republiky k EÚ 15– predpoklad úspešného členstva Slovenska v európskej menovej únii. Ekonomický časopis, 52, č. 9, 2004, s. 1064-1079

[33] THANASSOULIS, E.: Assessing police forces in England and Wales using data envelopment analysis, European Journal of Operational Research, Volume 87, Issue 3, 21 December 1995, Pages 641-657

Adresy autorov:

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

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

Dagmar Markechová, Doc. RNDr., CSc.
Katedra Matematiky UKF
Tr. A. Hlinku 1
949 01 Nitra1
atirpakova@ukf.sk

Jozef Stieranka, Doc. Ing., PhD.
Katedra kriminálnej polície
Akadémia policajného zboru
Sklabinská 1
835 17 Bratislava 35
jozef.stieranka@minv.sk

Modely počtu poistných plnení
Models of the Numbers of Claims

Marta Urbaníková

Abstract: The contribution deals with the possibility of modeling the number of claims in individual insurance products by the means of probability distributions and their characteristics. The probability distribution allows to describe the nature of extensive insurance portfolio as well as test their properties by the help of simple mathematic relations.

Key words: Insurance, Probability distributions, numbers of claims

Kľúčové slová: poistenie, pravdepodobnostné rozdelenia, počet poistných plnení.

1. Úvod

Poistenie možno považovať za nástroj na ochranu proti rizikám. Každá poistná udalosť má charakter náhodnej udalosti, obyčajne zriedkavej a veľmi málo pravdepodobnej, ale s mimoriadne závažnými dopadmi pre poisteného v prípade jej vzniku. Poistením si poistený zabezpečuje právo na výplatu poistného plnenia v prípade vzniku poistnej udalosti. Poistenie je poskytované za úplatu – poistné.

Pri výpočte poistného môžu byť použité dva prístupy: deterministický a stochastický. Pri deterministickom prístupe sa vychádza zo štatistických podkladov pre jednotlivé tarifné skupiny a kalendárne roky, na základe ktorých sa získavajú štatistické ukazovatele – priemerné poistné plnenie, priemerná poistná suma, priemerná škoda, škodová frekvencia, škodový stupeň. Na základe týchto údajov sa potom kalkuluje poistné. Druhý prístup je stochastický. Ide o pravdepodobnostné modely poistných javov. Tieto modely umožňujú nahradiť nedostatočný počet dát (v prípade nedostatočnej predchádzajúcej evidencii dát, v prípade nových produktov, v začínajúcich poisťovniach a pod.). Ďalšou výhodou tohto prístupu je, že často pomocou jednoduchých matematických vzťahov s malým počtom parametrov sa dá opísať chovanie rozsiahlych poistných kmeňov a dajú sa štatisticky testovať ich vlastnosti .

2. Modely počtu škôd

Ide o modely vychádzajúce z pravdepodobnostného rozdelenia náhodnej premennej n, ktorá označuje počet poistných škôd, teda počet poistných udalostí, resp. počet poistných plnení obvykle na jednu poistnú zmluvu počas jedného roka. Náhodná premenná n väčšinou môže nadobúdať hodnoty 0, 1, 2, ... s určitými pravdepodobnosťami. Počet poistných plnení n má najčastejšie niektoré z nasledujúcich rozdelení:

- Binomické
- Poissonovo
- Negatívne binomické rozdelenie
- Zmiešané Poissonovo rozdelenie.

1. **Binomické rozdelenie** $n \sim \mathrm{Bi}(K, p)$

Je to dvoj parametrické rozdelenie s prirodzeným K a $0 < p < 1$. Náhodná premenná n predstavuje počet úspechov v K nezávislých pokusoch s pravdepodobnosťou nastatia úspechu p a pravdepodobnosťou neúspechu 1-p. Potom

$$P(n=k) = \binom{K}{k} p^k (1-p)^{K-k}, \qquad k = 0,1,...,K$$

$$E(n) = Kp \tag{1}$$

$$D(n) = Kp(1-p) \tag{2}$$

Z (1) a (2) vyplýva, že $E(n) > D(n)$

2. **Poissonovo rozdelenie** $n \sim Po(\lambda)$

Toto rozdelenie je jedno parametrické rozdelenie s parametrom $\lambda > 0$. Vzniká ako limitný prípad binomického rozdelenia $n_k \sim \mathrm{Bi}(K, p_K)$ pre $K \to \infty$ a $p_k \to 0$, pričom stredná hodnota je $E(n_K) = Kp_K = \lambda$ Poissonovo rozdelenie je rozdelenie pravdepodobnosti riedkych javov. Náhodná premenná n predstavuje počet úspechov v sérii veľkého počtu nezávislých pokusov s malou pravdepodobnosťou nastatia úspechu . Pravdepodobnostná funkcia má tvar

$$P(n=k) = e^{-\lambda}\frac{\lambda^k}{k!}, \; k = 0,1,2,...,$$

$$E(n) = \lambda$$

$$D(n) = \lambda$$

Nakoľko $E(n) = \lambda$, hodnota parametra λ v prípade Poissonovho rozdelenia sa rovná škodovej frekvencii q_1, teda $n \sim Po(q_1)$.

3. **Negatívne binomické rozdelenie (Pólyovo rozdelenie)** $n \sim NB(\alpha, p)$

je dvoj parametrické rozdelenie s $\alpha > 0$ a $0 < p < 1$. V prípade, že α je prirodzené číslo, náhodná premenná n predstavuje počet neúspechov pred α-tým úspechom v nezávislých pokusoch s pravdepodobnosťou nastatia úspech p a pravdepodobnosťou neúspechu 1-p . Potom

$$P(n=k) = \binom{\alpha+k-1}{k} p^{\alpha}(1-p)^k, \; k = 0,1,2,...,$$

$$E(n) = \frac{\alpha(1-p)}{p}$$

$$D(n) = \frac{\alpha(1-p)}{p^2}$$

Pomocou Poissonovho a negatívneho binomického rozdelenia môžeme modelovať počet škôd na jednu poistnú zmluvu počas jedného roka aj pri najväčšom poistnom segmente pri poistení zodpovednosti za škodu spôsobenú prevádzkou motorového vozidla

4. Zmiešané Poissonovo rozdelenie

je Poissonovo rozdelenie, ktorého parameter λ je náhodný s pravdepodobnostnou hustotou $f(\lambda)$:

$$P(n=k)=\int_0^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}f(\lambda)d\lambda,\ k=0,1,2,...,$$

Zmiešané Poissonovo rozdelenie sa používa pre poistné kmene s heterogénnymi rizikami, pričom pri poistných zmlúvách s malým (resp. veľkým) rizikom nadobúda parameter λ malé (resp. veľké) hodnoty. Týmto rozdelením sa napr. môže riadiť počet *n* lesných požiarov v danom regióne počas letných mesiacov.

Výber vhodného rozdelenia počtu poistných plnení (škôd) možno uskutočniť na základe vzťahu medzi strednou hodnotou a disperziou náhodnej premennej nasledovne:

- Ak $E(n) > \mathrm{D}(n)$, odporúča sa použiť Binomické rozdelenie
- Ak $E(n) = D(n)$ odporúča sa použiť Poissonovo rozdelenie
- Ak $E(n) < D(n)$ odporúča sa použiť Negatívne binomické rozdelenie

3. Modely počtu škôd portfólia zmlúv

Doteraz sme sa zaoberali modelovaním počtu škôd počas jedného roka v prípade jednej poistnej zmluvy. Teraz predpokladajme, že poisťovňa má uzatvorených *N* nezávislých poistných zmlúv. Nech náhodná premenná *n* označuje počet poistných plnení počas jedného roka v *N* navzájom nezávislých poistných zmluvách. Potom

$$n=n_1+n_2+n_3+...+n_N,$$

kde n_i je počet poistných plnení počas jedného roku v *i*-tej zmluve. Z vlastností uvažovaných pravdepodobnostných rozdelení vyplýva:

- Ak $n_i \sim Bi(K_i,p)$, tak $n \sim Bi(K_1+...+K_N,p)$
- Ak $n_i \sim Po(\lambda_i)$, tak $n \sim Po(\lambda_1+...+\lambda_N)$
- Ak $n_i \sim NB(\alpha_i,p)$, tak $n \sim NB(\alpha_1+...+\alpha_N,p)$

V prípade homogénneho portfólia s rovnakou škodnou frekvenciou q_1, platí v prípade Poissonovho rozdelenia

- Ak $n_i \sim Po(q_1)$ tak $n \sim Po(Nq_1)$.

4. Asymptotické aproximácie počtu škôd

Na základe Centrálnej limitnej vety môžeme pri väčších poistných kmeňoch využiť za splnenia istých podmienok aproximácie Normálnym rozdelením. Veľmi často sa využíva aproximácia Poissonovho rozdelenia Normálnym rozdelením. Táto aproximácia je veľmi výhodná, lebo podstatne zjednoduší výpočty pravdepodobnosti intervalov možných hodnôt.

5. Záver

Pri výpočtoch poistného hrá dôležitú úlohu modelovanie počtu a výšky škôd pomocou pravdepodobnostných rozdelení. Samotná kalkulácia rizikového poistného sa veľmi často robí za pomoci stochastických modelov. Výhoda je v tom, že stačí definovať iba základné závislosti, parametre rozdelení a model už urobí všetko sám.

Je nutné si uvedomiť, že kalkulácia je proces, nie úloha. Počíta sa dovtedy, kým aj riziko aj cena nie sú optimálne a konkurencieschopné. Aj kalkulácia bezpečnostnej prirážky vychádza práve zo stochastického modelu. Na základe výsledkov napr. 5000 simulácií poisťovňa môže ľahko vidieť, kde je 95 – percentil a vie tým prispôsobiť cenu.

6. Literatúra

[1] CIPRA, T. 1999. Poistná matematika teorie praxe a praxe. Praha Ekopress, 1999. 398 s. ISBN 80-86119-17-3.

[2] PACAKOVÁ, V. 2000. Aplikovaná poistná štatistika. Bratislava ELITA, 2000. 248 s. ISBN 80-8044-073-5.

[3] URBANÍKOVÁ, M – VACULÍKOVÁ, Ľ. 2006. Aktuárska matematika. Bratislava STU, 2006. 186 s. ISBN 80-227-2442-4.

Adresa autora :

Marta Urbaníková, doc. RNDr. CSc.
Levická 72
949 01 Nitra
murbanikova@ukf.sk

Stanovenie rozsahu náhodného výberu pri testovaní hypotéz o strednej hodnote, ak poznáme/ nepoznáme rozptyl základného súboru
Determination of Sample Size for Hypothesis Tests of the Mean Value, if the Case of Known / Not Known Variance of Population

Božena Viktorínová

Abstract: The article deals with determining the sample size for hypothesis tests of the mean value and the possibility of reiterating the sample size for more exact results.

Key words: Testing, hypothesis, Sample size, Iteration

Kľúčové slová: testovanie, hypotéza, rozsah výberu, iterácia

1. Úvod

Budeme sa zaoberať stanovením rozsahu náhodného výberu zo základného súboru ak závery o tomto rozsahu môžeme iteráciou meniť. Napríklad [1] ak bude treba stanoviť rozsah výberu automobilových pneumatík rovnakého druhu, ktorých opotrebovanie po X kilometroch je rôzne, a z pôvodnej výšky dezénu zostáva v priemere Y milimetrov. Potom samozrejme budeme testovať hypotézu o priemernej výške dezénu celého základného súboru po X kilometroch. V takýchto prípadoch po stanovení rozsahu výberu z rovnice (1) a po výpočte rozsahu výberu z rovnice (2), môžeme výsledok rovnice (2) iterovať (znovu vypočítať) pomocou rovnice (2), a tak ho upresniť.

2. Popis uvedenej problematiky,

čiže stanovenie rozsahu výberu, ak budeme testovať strednú hodnotu základného súboru. Aby sme mohli vyhodnotiť hypotézu o strednej hodnote základného súboru na základe testovacieho kritéria, musíme mať k dispozícii riziko dodávateľa α, riziko odberateľa β a prijateľnú nespoľahlivosť δ. Problém riešime najprv ak je známa štandardná odchýlka základného súboru σ a potom ak štandardnú odchýlku základného súboru nepoznáme. Ak ju poznáme, potom δ môžeme vyjadriť pomocou vzťahu $\delta = 0{,}5.\sigma$, kde číslo 0,5 vyjadruje nami prípustný „podiel“ zo štandardnej odchýlky.

Potom rozsah výberu určíme z rovnice:

$$n = (U_\alpha + U_\beta)^2 . \frac{\sigma^2}{\delta^2} \qquad (1)$$

V tejto rovnici U_α a U_β určíme z Tab. B, [1], str.696, pre jednostranný test, kde U_α a U_β sú kritické hodnoty normálneho rozdelenia.

Tabuľka 1:(Tab .B)Kritické hodnoty normálneho rozdelenia pre jednostranný test, napr.: $\mu \le \bar{x} + \frac{U_\alpha . \sigma}{\sqrt{n}}$ ***, ak rozptyl*** σ^2 ***je známy:***

α alebo β	U	α alebo β	U
0,001	3,090	0,100	1,282
0,005	2,576	0,150	1,036
0,010	2,326	0,200	0,842
0,015	2,170	0,300	0,524
0,020	2,054	0,400	0,253
0,025	1,960	0,500	0,000
0,050	1,645	0,600	- 0,053

Ak ale bude alternatívna hypotéza obojstranná ($\mu\langle$ alebo $\rangle$ kritérium) potom U_α určíme z Tab. C.

Tabuľka 2:(Tab .C) Kritické hodnoty normálneho rozdelenia pre obojstranný test, ak rozptyl σ^2 je známy:

iba α	U	iba α	U
0,001	3,291	0,100	1,645
0,005	2,807	0,150	1,440
0,010	2,576	0,200	1,282
0,015	2,432	0,300	1,036
0,020	2,326	0,400	0,842
0,025	2,241	0,500	0,675
0,050	1,960	0,600	0,524

Ak štandardná odchýlka nie je známa, rozsah výberu určíme pomocou vzorca:

$$n = (t_\alpha + t_\beta)^2 \cdot \frac{s^2}{\delta^2} \qquad (2)$$

V tejto rovnici t_α a t_β určíme z Tab. D, ktorú tu neuvádzame ,ale nachádza sa napr. v [1], str.697, čo sú vlastne kvantily Studentovho rozdelenia pre počet stupňov voľnosti $\upsilon = n - 1$ a kritickú oblasť α a β a to v prípade ak sa jedná o jednostrannú alternatívnu hypotézu, t. zn. $\mu\langle$ kritérium. Ak alternatívna hypotéza je obojstranná, t. zn., $\mu\langle$ alebo $\rangle$ kritérium, potom t_α určíme z Tab. E, [1], str.698. Je to tabuľka kritických hodnôt t – rozdelenia pri obojstrannom teste.

3. Príklad

na stanovenie rozsahu výberu pri testovaní strednej hodnoty základného súboru **,** ak štandardná odchýlka je známa.

Opotrebovanie dezénu pneumatík rovnakého druhu po 40 000 km je v priemere 6 mm. Treba určiť rozsah výberu, ktorý je potrebný na to, aby sme overili toto kritérium, ak máme k dispozícii tieto údaje:

$\alpha = 0{,}1$ (riziko dodávateľa)

$\beta = 0{,}05$ (riziko odberateľa)

$\delta = 0{,}5.\sigma$ (prijateľná nespoľahlivosť, čo je polovica štandardnej odchýlky)

Aby sme mohli dosadiť do rovnice (1) potrebné údaje, nájdeme v Tab. B hodnotu $U_{0,1} = 1{,}282$ a $U_{0,05} = 1{,}645$ potom

$$n = (1{,}282 + 1{,}645)^2 . \frac{\sigma^2}{0{,}25.\sigma^2} = \frac{8{,}567}{0{,}25} = 34{,}268 \qquad (3)$$

Ak (napriek pravidlu o zaokrúhľovaní) zaokrúhlime toto číslo smerom nahor $[1]$, dostaneme rozsah výberu $n = 35$.

V prípade, ak štandardná odchýlka nie je známa, tento rozsah výberu určíme z rovnice (2), pričom počet stupňov voľnosti pre t – rozdelenie určíme ako $\upsilon = n - 1 = 35 - 1 = 34$, teda z výsledku rovnice (3). Keďže kritické hodnoty $t_{0.1,34}$ a $t_{0.05,34}$ sa v Tab. D nenachádzajú, dostaneme ich interpoláciou nasledovne: v Tab. D sme našli $t_{0.1,30} = 1{,}310$ a $t_{0.1,40} = 1{,}303$. Rozdiel týchto dvoch tabuľkových hodnôt je 0,007, čo zodpovedá 10 dielikom. Potom štyrom dielikom zodpovedá hodnota 0,0028. Ak odpočítame od 1,310 – 0,0028, dostaneme hodnotu 1,3072 čo je aj hodnotou $t_{0.1,34} = 1{,}3072$. Podobne interpoláciou medzi hodnotami $t_{0.05,30} = 1{,}697$ a $t_{0.05,40} = 1{,}684$ nájdeme, že $t_{0.05,34} = 1{,}6918$. Obidva výsledky získané interpoláciou dosadíme do rovnice (2):

$$n = (1{,}692 + 1{,}307)^2 . \frac{s^2}{0{,}25.s^2} = \frac{8{,}994}{0{,}25} = 35{,}976 \qquad (4)$$

Zaokrúhliac tento výsledok bude rozsah výberu rovný 36. Táto rovnica (4) a teda aj (2) by mohla byť znovu prepočítaná pre $\upsilon = 36 - 1 = 35$ stupňov voľnosti, a takto by sme mohli pokračovať stále ďalej, teda iterovať rozsah výberu.

Zistilo sa, že pri výpočtoch výsledok veľmi ovplyvňujú nízke hodnoty parametrov α, β a δ, čo vedie k veľkým rozsahom výberov.

4. Záver

Pri testovaní hypotéz o strednej hodnote, rozsah výberu určujeme podľa rovnice (1) resp. (2), pričom rovnica (2) závisí od výsledku rovnice (1). Tiež si myslíme, že zaokrúhľovanie výsledku výpočtu podľa (3) tak, ako je to uvedené v príklade , neovplyvní významne výpočet podľa rovnice (2), pretože ak by sme výsledok 34,268 zaokrúhlili smerom nadol, bolo by $\upsilon = 34 - 1 = 33$ a potom $t_{0.1,33} = 1{,}308$, $t_{0.05,33} = 1{,}693$ a n podľa (2) by bolo rovné 36,02 , čo sa nelíši od výsledku rovnice (4). Táto teória práve pre jej možnosť iterácie (a upresňovania) rozsahu výberu sa vidí byť dosť zaujímavá, aj keď nie nevyhnutná.

5. Literatúra

[1] FORREST, W., BREYFOGLE,III.: Implementing six sigma, John Wiley & Sons, inc., Toronto 2003, ISBN 978 – 0 – 471 – 265 – 72 – 6, str. 295 – 296.

[2] TEREK,M., HRNČIAROVÁ, Ľ.,: Štatistické riedenie kvality, Bratislava, IURA EDITION, 2004, ISBN 80 – 89047 – 97 – 1.

[3] PAŽITNÁ,M., LABUDOVÁ, V.: Metódy štatistického porovnávania, Bratislava, EKONÓM, 2007, ISBN 978 – 80 – 225 – 2285 – 4.

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

[5] ŠOLTÉS,E., ŠOLTÉSOVÁ,T.: Overovanie predpokladu normality. In.:Zborník príspevkov z 1.vedeckého seminára :Štatistické metódy v ekonómii, Bratislava, Ekonóm, 2007, ISBN 978 – 80 – 225 – 2423 – 0.

Adresa autora:

Božena Viktorínová, Mgr., CSc.
Dolnozemská 1
852 35 Bratislava
Ekonomická univerzita, FHI, KŠ
viktorin@dec.euba.sk

Tento príspevok vznikol s prispením grantovej agentúry VEGA, v rámci projektu číslo 1/0437/08:Kvantitatívne metódy v stratégii šesť sigma

Štatistické analýzy výsledkov medzilaboratórnych skúšok
The Statistical Analysis of the Results of Collaborative Trials

Marta Vrábelová, Július Jenis

Abstract: The statistical analysis of collaborative trial results of the measurements of the somatic cell number in milk are described in this paper.

Key words: collaborative trials, the somatic cell number in milk, assigned value, repeatability, reproducibility.

Kľúčové slová: počet somatických buniek v mlieku, referenčná hodnota, opakovateľnosť, reprodukovateľnosť.

1. Úvod

Medzilaboratórne skúšky testovania rôznych meracích prístrojov sa uskutočňujú za účelom kontroly a kvality laboratórnej práce, ktorá je dôležitá hlavne z hľadiska zabezpečenia kvality a bezpečnosti potravín. Tieto skúšky vykonávajú štátne veterinárne a potravinové ústavy, prostredníctvom ktorých získavajú laboratóriá akreditáciu. Skúšky sa vykonávajú podľa normy. Dôležitou potravinovou zložkou je mlieko. Kvalita mlieka je daná aj počtom somatických buniek. Skúšky meraní počtu somatických buniek v mlieku popisuje norma STN EN ISO 13366-3. My sa v tomto článku zameriame len na popis štatistických analýz výsledkov týchto skúšok a ich výpočet v Exceli.

2. Vzorky a merania

Každé laboratórium zapojené do skúšky obdrží 40 vzoriek mlieka, pričom ide o 10 rôznych vzoriek, z ktorých každá je rozdelená do 4 vzoriek. Identitu týchto očíslovaných vzoriek pozná len skúšobné laboratórium. Laboratóriá zmerajú každú obdržanú vzorku 4-krát, teda každé laboratórium urobí 160 meraní, pričom máme 16 meraní každej z 10-tich vzoriek. Počty buniek v týchto vzorkách sa pohybujú od 200 000 do 800 000 buniek v jednom mililitri.

3. Štatistické spracovanie výsledkov

Štatistické analýzy tiež predpisuje norma, ktorá pripúšťa aj použitie logaritmov nameraných hodnôt. Tento predpísaný postup spracovania výsledkov je zvyčajne pre pracovníkov skúšobných laboratórií nedostatočný a potrebujú pomoc štatistika. Takto vznikol (amatérsky vyrobený) program. v Exeli, pre daný počet laboratórií, ktorý po vložení aktuálnych údajov vypočíta všetky potrebné charakteristiky a tabuľky. Tento program teraz popíšeme.

4. Postup vyhodnotenia združených testov pre stanovenie počtu somatických buniek v mlieku

Pri vyhodnocovaní medzilaboratórnych skúšok sme sa pridŕžali postupu uvedeného v STN EN ISO 13366 – 3 a protokolu zostaveného Wiliamom Horwitzom [2].

(1) Vypočítali sme priemery stanovení počtu buniek jednotlivých laboratórií a celkový priemer pre každú z 10 vzoriek.

(2) Pre každú vzorku a každé laboratórium sme vypočítali hranicu opakovateľnosti podľa vzorca

$$r = 2{,}8\sqrt{(s^2 + s_s^2)} = 2{,}8s_r,$$

s^2 je reziduálny rozptyl - rozptyl vo vnútri 4 skupín danej vzorky,

$s_s^2 = \frac{s_A^2 - s^2}{4}$, kde s_A^2 je rozptyl medzi 4 skupinami danej vzorky.

s_r je smerodajná odchýlka opakovateľnosti.

Vypočítali sme priemernú hodnotu $\bar{r}$ cez všetky laboratória pre každú vzorku. Ak pre niektorú vzorku v niektorom laboratóriu platilo $r > 1{,}15\ \bar{r}$, priemer tejto vzorky bol vynechaný pri výpočte korigovaného celkového priemeru pre danú vzorku.

(3) Pre každé laboratórium sme vypočítali koeficienty regresnej priamky $y = a + bx$ závislosti priemerných hodnôt laboratória od celkových priemerných hodnôt 10 vzoriek. Pre každé laboratórium *i* sme vypočítali maximálne vychýlenie podľa vzorca

$\max\{| a + b\bar{x}_{ij} - \bar{x}_j |, j = 1{,}2{,}...{,}10\}$, kde $\bar{x}_{ij}$ je priemer *j* - tej vzorky v *i* - tom laboratóriu a $\bar{x}_j$ je celkový priemer *j* – tej vzorky.

Vypočítali sme priemerné maximálne vychýlenie cez všetky laboratóriá. Ak maximálne vychýlenie niektorého laboratória presiahlo 1,15 násobok priemerného maximálneho vychýlenia, toto laboratórium bolo vynechané pri výpočte korigovaných celkových priemerov.

(4) S prihliadnutím na bod 2 a bod 3 boli vypočítané korigované celkové priemery – referenčné stredné hodnoty.

(5) Pre každé laboratórium bola vypočítaná nová regresia priemerných hodnôt vzoriek laboratória k referenčným stredným hodnotám a nové maximálne vychýlenie

$\max\{a + b\bar{x}_{ij} - \bar{x}_j, j = 1{,}2{,}...{,}10\}$ - korigovaný BIAS.

Vypočítaný bol tiež štandardný BIAS, čo je odchýlka hodnoty regresnej priamky na úrovni 500 (.1000) somatických buniek ($a + b$.500 – 500).

(6) Pre každú vzorku sme vypočítali hranicu reprodukovateľnosti cez všetky laboratóriá podľa vzorca

$$R = 2{,}8\sqrt{s_r^2 + s_L^2} = 2{,}8s_R, \text{ kde } s_L^2 = \frac{s_B^2 - s^2}{7}.$$

Teraz s^2 je rozptyl vo vnútri laboratórií - reziduálny rozptyl a s_B^2 je medzilaboratórny rozptyl. s_R je smerodajná odchýlka reprodukovateľnosti. Uvádzame tiež variačný koeficient - relatívnu smerodajnú odchýlku opakovateľnosti v % a relatívnu smerodajnú odchýlku reprodukovateľnosti v %

$$\text{RSDr} = \frac{s_r}{\bar{x}}100, \ \text{RSDR} = \frac{s_R}{\bar{x}}100, \text{ kde } \bar{x} \text{ je laboratórny priemer za danú vzorku.}$$

Tieto výsledky prezentujeme len pre vzorku s nízkym počtom somatických buniek (PSB) a vzorku s vysokým počtom somatických buniek. Pre tieto vzorky sa za rozumné hodnoty variačných koeficientov považujú hodnoty:

Pre nízky počet somatických buniek (100 000 - 200 000),

RSDr: 5% -10%, RSDR: 10% -20%.

Pre vysoký počet somatických buniek (400 000 – 500 000),
RSDr: 4% - 5%, RSDR: 10% - 12%.
(7) Vypočítali sme smerodajnú odchýlku opakovateľnosti zo všetkých 40 vzoriek pre každé laboratórium ako v bode 2.
(8) Výsledky každého laboratória sme zobrazili graficky. Graf (obrázok 1) obsahuje priamku $y = x$, regresnú priamku $y = a + bx$ a pre každú vzorku aj maximálnu a minimálnu zistenú hodnotu spojenú úsečkou.

Pri počítaní korigovaných priemerov treba zaistiť, aby sa ich hodnoty počítali z priemerov minimálne 5 *laboratórií. Pri počte* 7 *laboratórií by sa pri počítaní celkových priemerov malo úplne vynechať najviac* 1 *laboratórium (*2/9 *z celkového počtu).*

4. Vyhodnotenie

Výsledky získané pomocou štatistických analýz sa rozposielajú skúšaným laboratóriám vo forme piatich tabuliek obsahujúcich:

- priemerné hodnoty počtu somatických buniek (.1000 v 1 ml) pre každé laboratórium a každú vzorku mlieka s uvedením referenčných hodnôt vypočítaných v bode (4);
- opakovateľnosť (.1000) a reprodukovateľnosť (.1000) stanovenia na nízkej hladine PSB (bod (6));
- opakovateľnosť (.1000) a reprodukovateľnosť (.1000) stanovenia na vysokej hladine PSB (bod (6));
- smerodajné odchýlky opakovateľnosti vypočítané v bode (7) zo všetkých vzoriek pre každé laboratórium usporiadané podľa veľkosti (tabuľka 1);
- smernice a posunutia regresnej priamky, korigované a štandardné odklony (BIAS) pre každé laboratórium vypočítané v bode (5) (tabuľka 2)

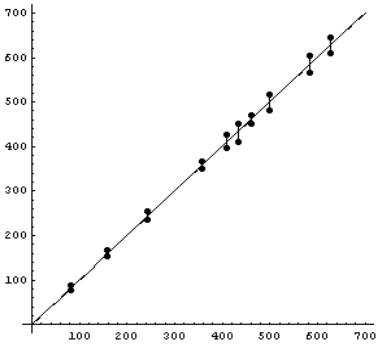
a grafu, ktorý obsahuje individuálne počty daného laboratória proti referenčným stredným hodnotám (obrázok 1).

Tabuľka 1: Smerodajná odchýlka opakovateľnosti vypočítaná zo všetkých vzoriek

	s_r so všetkých vzoriek
Laboratórium 1	54,4
Laboratórium 2	54,34
Laboratórium 3	54,57
Laboratórium 4	54,54
Laboratórium 5	54,56
Laboratórium 6	54,31
Laboratórium 7	54,69

Tabuľka 2: Smernica a posunutie regresnej priamky, korigovaný a štandardný BIAS

	Smernica	Posunutie	Kor. BIAS	Štand. BIAS
Laboratórium 1	0,9999	0,303	0,29	0,23
Laboratórium 2	0,9986	0,541	0,43	-0,14
Laboratórium 3	1,0009	-1,009	-0,94	-0,57
Laboratórium 4	1,0028	0,239	2,02	1,65
Laboratórium 5	0,9981	-0,988	-2,16	-1,92
Laboratórium 6	0,9985	0,33	-0,59	-0,4
Laboratórium 7	1,0039	-2,129	-1,81	-0,18



Obrázok 1: Individuálne počty buniek proti referenčným stredným hodnotám

5. Záver

K spracovávaniu výsledkov laboratórnych skúšok existuje viacero protokolov, spomenieme aspoň technickú správu Thompsona a Wooda [3] a tiež množstvo článkov, napr. [1], [4] a ďalších dokumentov vydávaných spoločnosťou *Royal Society of Chemistry*, dostupných na www stránke [5]. Pripravujeme profesionálnu verziu vyššie uvedeného programu.

6. Literatúra

[1] BURKE, S. Missing Values, Outliers, Robust Statistics & Non-parametric Method. EC+GC Europe Online Supplement, Esarda Bulletin, No. 28, s. 19-24, dostupné na stránke: http://www.lcgceurope.com/lcgceurope/data/articlestandard/lcgceurope/502001/4509/article.pdf

[2] HORWITZ, W. 1995. Protocol for the Design, Conduct and Interpretation of Method-Performance Studies. Pure & Appl. Chem., Vol. 67, No. 2, s. 331-343, 1995.

[3] THOMPSON, M. – WOOD, R. 1993. The International Harmonized Protocol for The Proficiency Testing of (Chemical) Analztical Laboratories. Pure & Appl. Chem., Vol. 65, No. 9, s. 2123-2144, 1993.

[4] THOMPSON, M. 2000. Recent Trends in Interlaboratory Precision at ppb and subppb Concentrations in Relation to Fitness for Purpose Criteria in Proficiency Testing. In: Analyst, No. 125, s. 385-386, 2000.

[5] www.rsc.org/lap/rsccom/amc/amc_index.htm

Adresy autorov:

Marta Vrábelová, doc. RNDr., CSc.
KM FPV UKF, Tr. A. Hlinku 1
949 74 Nitra
mvrabelova@ukf.sk

Július Jenis, RNDr.
CIKT UKF, Tr. A. Hlinku 1
949 74 Nitra
jjenis@ukf.sk

Využitie logistickej regresie pri odhadovaní hraníc subjektívnej chudoby
Application of Logistic Regression in Subjective Poverty Lines Estimation

Tomáš Želinský

Abstract: The paper analyzes perception of subjective poverty in Slovakia. The analysis is based on so called "discrete information" approach. Logistic regression is then applied as a tool for subjective poverty analysis and estimation of subjective poverty lines. Using the EU SILC microdata, three alternatives are compared (in terms of who can be considered poor). Two of the three alternatives do not match the empirical data very well, i. e. they overestimate or underestimate the level of subjective poverty. The estimated proportions of poor are then inadequate and should not be considered.
According to the adequate alternative approximately 20 per cent of Slovak population can be considered subjectively poor (in terms of subjective perception of poverty in relation to equivalent disposable income).

Keywords: subjective poverty, poverty line, logistic regression, EU SILC.

Kľúčové slová: subjektívna chudoba, hranica chudoby, logistická regresia, EU SILC.

1. Úvod

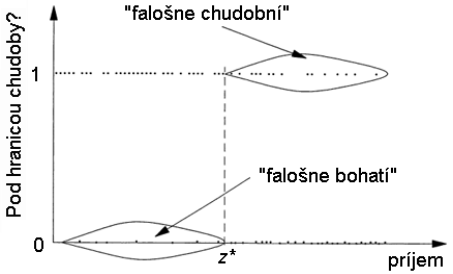
Príspevok nadväzuje na článok o prístupe k meraniu subjektívnej chudoby založenom na porovnávaní distribučných funkcií rozdelenia príjmov dvoch skupín obyvateľov – tých, ktorí sa považujú za chudobných a tých ostatných [1].

Cieľom príspevku je ponúknuť ďalší nástroj na odhad subjektívnej chudoby – logistickú regresiu, resp. regresný model s kvalitatívnou vysvetľovanou premennou (v našom prípadne binárnou premennou).

2. Určenie hraníc subjektívnej chudoby

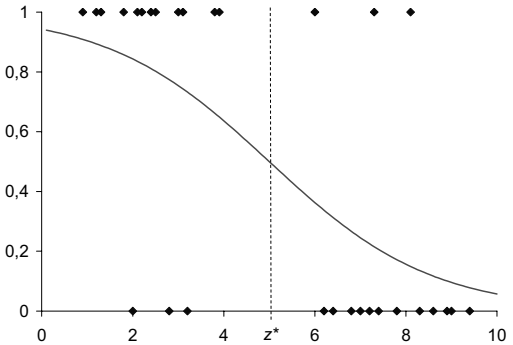
Subjektívny prístup k definovaniu chudoby vyjadruje, že hranice chudoby sú inherentnými subjektívnymi úsudkami ľudí o tom, čo považujú za sociálne akceptovateľný minimálny životný štandard v určitej spoločnosti [2]. Prístup vychádza z predpokladu, že podmienky, v ktorých sa jednotlivec nachádza v porovnaní k okolnostiam ostatných členov referenčnej skupiny tak, ako ich vníma, ovplyvňujú vnímanie jeho osobného blahobytu relatívne k ostatným členom referenčnej skupiny [3].

Pri určovaní hranice subjektívnej chudoby budeme opäť vychádzať z prístupu založenom na tzv. „diskrétnom položení otázky" [4]. Pripomenieme, že diskrétnosť otázky spočíva v tom, že respondenti nemajú stanoviť absolútnu hodnotu príjmu, ktorý považujú za minimálne požadovaný. Majú v odpovedi uviesť, či vzhľadom na svoj aktuálny príjem sa cítia byť chudobní (hodnota „1"), alebo nie (hodnota „0"). Následne je žiaduce odhadnúť hodnotu subjektívnej hranice z^*. Hodnota z^* je určená ako hodnota, ktorej zodpovedá najvyššia pravdepodobnosť, že odpovede respondentov o ich statuse chudoby korešpondujú so statusom, ktorý by sme zistili porovnaním hodnoty z^* s ich príjmom. Situácia je graficky zachytená na obr. 1.



Obrázok 1: Odhad subjektívnej hranice chudoby
Zdroj: [4, s. 125]

V procese odhadu chudoby možno využiť logistickú regresiu – hľadáme logistickú krivku (resp. distribučnú funkciu logistického rozdelenia), ktorá najlepšie vyhovuje empirickým hodnotám. Graficky je situácia znázornená na obr. 2.



Obrázok 2: Využitie logistickej regresie pri odhade subjektívnej chudoby
Zdroj: vlastný graf

3. Metódy

3.1 Model

Majme dvojrozmerné pozorovania $[X_i; Y_i]$, kde X_i je aktuálny príjem i-tej osoby a Y_i je binárna premenná popisujúca subjektívne vnímanie pocitu chudoby i-tej osoby, a teda môže nadobúdať hodnotu 0 (nie je subjektívny pocit chudoby) alebo 1 (je subjektívny pocit chudoby).

Nech $P(y_i = 1 \mid x_i)$ je pravdepodobnosť, že i-tú osobu budeme považovať za chudobnú za predpokladu daného príjmu tejto osoby a analogicky $P(y_i = 0 \mid x_i)$ je pravdepodobnosť, že i-tú osobu nebudeme považovať za chudobnú pri danom príjme x_i.

Uvažujúc *logitový model* (napr. [5]) máme:

$$P(y_i = 1 \mid x_i, b_1, b_2) = \frac{e^{b_1 + b_2 x_i}}{1 + e^{b_1 + b_2 x_i}}, \tag{1}$$

$$P(y_i = 0 \mid x_i, b_1, b_2) = 1 - \frac{e^{b_1 + b_2 x_i}}{1 + e^{b_1 + b_2 x_i}}. \tag{2}$$

3.2 Odhad hranice subjektívnej chudoby

Za hranicu chudoby z^* budeme považovať hodnotu príjmu, ktorá zodpovedá hodnote 0,5 distribučnej funkcie odhadnutého logistického rozdelenia. Vzhľadom na to, že logistické rozdelenie je symetrické, daným bodom je inflexný bod, tzn.

$$z^* = -\frac{b_1}{b_2} \tag{3}$$

Iným prípadom by bolo použitie napríklad modelu GOMPIT založeného na Gumbelovom rozdelení, ktoré má nesymetrickú distribučnú funkciu, kedy by sa tieto dva body nemuseli zhodovať.

3.3 Popis vstupných údajov

Vstupnými údajmi pre analýzu sú mikroúdaje zisťovania EU SILC 2006 [6]. Pre detaily týkajúce sa popisu vzorky a zberu údajov, pozri [7].

Za *príjem* budeme v modeli pokladať *ekvivalentný disponibilný príjem domácnosti*, t. j. podiel celkového ročného disponibilného príjmu domácnosti a ekvivalentnej veľkosti domácnosti [7].

Domácnosťou rozumieme hospodáriacu domácnosť, t. j. súkromnú domácnosť tvorenú osobami v byte, ktoré spoločne žijú a spoločne hospodária, vrátane spoločného zabezpečovania životných potrieb. Za znak spoločného hospodárenia sa považuje spoločná úhrada základných výdavkov domácnosti (strava, úhrada nákladov za bývanie, elektrina, plyn a pod.) [7]

Vnímanie subjektívnej chudoby domácnosťou hodnotíme na základe otázky v dotazníku: *„Domácnosť môže mať rôzne zdroje príjmu a viac ako jeden člen domácnosti môže prispievať do celkového príjmu. Ak sa zamyslíte nad celkovým mesačnými príjmom Vašej domácnosti, s akými ťažkosťami je Vaša domácnosť schopná splatiť zvyčajné výdavky?“*

1. s veľkými ťažkosťami,	*4. pomerne ľahko,*
2. s ťažkosťami,	*5. ľahko,*
3. s určitými ťažkosťami,	*6. veľmi ľahko.*

Vnímanie subjektívnej chudoby domácnosťou je binárnou premennou, preto je potrebné škálu odpovedí tejto otázky transformovať. Označme ako y_i odpoveď *i*-tej domácnosti, pričom $y_i \in \{1, 2, ..., 6\}$ a ako y_i^* transformovanú odpoveď tej istej domácnosti, kde $y_i^* \in \{0, 1\}$.

V snahe vyhnúť sa subjektivite pri tejto transformácii, uskutočníme tri porovnania vychádzajúce z troch transformácií premennej *Y*:

1. $\left(\forall y_i = 1 : y_i^* = 1\right) \& \left(\forall y_i > 1 : y_i^* = 0\right)$, (i)
2. $\left(\forall y_i \le 2 : y_i^* = 1\right) \& \left(\forall y_i > 2 : y_i^* = 0\right)$, (ii)
3. $\left(\forall y_i \le 3 : y_i^* = 1\right) \& \left(\forall y_i > 3 : y_i^* = 0\right)$. (iii)

(V prvom prípade považujeme za subjektívne chudobných len tých, ktorí označili možnosť „1“, v druhom prípade tých, ktorí označili možnosti „1“ alebo „2“ a v treťom prípade tých, ktorí označili „1“, „2“ alebo „3“.)

Vzhľadom na to, že v súbore sa nachádzalo niekoľko málo hodnôt príjmu, ktoré boli výrazne vysoké v porovnaní s ostatnými (ekvivalentný disponibilný príjem vyšší ako 400 000 Sk), dochádzalo k skresľovaniu výsledkov, a preto sme sa rozhodli tieto pozorovania zanedbať. Dokopy šlo o cca 60 pozorovaní.

Odhad parametrov modelu (1) resp. (2) sme uskutočnili v prostredí softvéru EViews 6.0.

4. Výsledky

4.1 Odhadnutá hranica chudoby

Ako sme naznačili, uskutočnili sme tri porovnania – zodpovedajúce trom prístupom k určeniu, koho považovať za subjektívne chudobného (pozri tab. 1).

Tabuľka 1: Odhad subjektívnej hranice v SR v roku 2006

Prístup	$P(y_i = 1 \mid x_i, b_1, b_2)$	**Hranica chudoby z*** *(ročná hodnota ekviv. disp. príjmu v SKK)*	**Podiel chudobného obyvateľstva** *(%)*
(i)	$\frac{e^{0,2987-0,0000193x_i}}{1+e^{0,2987-0,0000193x_i}}$	19 509,26	0,14
(ii)	$\frac{e^{1,2831-0,0000146x_i}}{1+e^{1,2831-0,0000146x_i}}$	87 608,74	19,92
(iii)	$\frac{e^{3,3678-0,0000128x_i}}{1+e^{3,3678-0,0000128x_i}}$	262 722,20	96,92

Zdroj: vlastné výpočty podľa údajov [6]

Z tab. 1 je zrejmé, že 1. prístup výrazne podhodnocuje a 3. prístup naopak výrazne nadhodnocuje podiel chudobného obyvateľstva. K záveru o nevhodnosti 1. a 3. modelu nás vedie aj porovnanie predpovedacích schopností modelov (tab. 2).

Tabuľka 2: Porovnanie predpovedacích schopností modelov

Model	*Absolútne početnosti binárnej premennej*			
	„0“		***„1“***	
	skutočné	*odhadnuté*	*skutočné*	*odhadnuté*
(i)	4360	5018	665	7
(ii)	3158	4024	1867	1001
(iii)	879	155	4146	4870

Zdroj: vlastné výpočty podľa údajov [6]

V prvom modeli je významne podhodnotený odhad hodnôt „1“, keď v skutočnosti ich bolo približne 95-krát viac ako odhadnutých modelom. Naopak v treťom modeli bol výrazne podhodnotený odhad hodnôt „0“ – v skutočnosti ich bolo približne 6-krát viac ako odhadnutých modelom.

Dochádzame preto k záveru, že za najvhodnejší možno považovať druhý model, na základe ktorého odhadujeme približne 20-percentný podiel obyvateľstva pociťujúceho (subjektívny) stav chudoby.

Uplatnený prístup nepriamo indikuje, že za subjektívne chudobných možno považovať tých, ktorí na stupnici so 6 kategóriami uviedli odpoveď „1“ alebo „2“, t. j. že domácnosť je schopná pokryť výdavky „*s veľkými ťažkosťami*“, prípadne „*s ťažkosťami*“. Uplatnením jednoduchého podielu týchto odpovedí by sme za subjektívne chudobných považovali 37,2% obyvateľstva.

5. Záver

Cieľom článku bolo demonštrovať využitie logistickej regresie (logitového modelu) pri odhade hraníc subjektívnej chudoby. Navrhovaný postup je možné použiť v analýze subjektívnej chudoby založenej na tzv. „diskrétnom položení otázky“ (t. j. respondenti neuvádzajú úroveň minimálne postačujúceho príjmu, ale na vopred stanovenej stupnici uvádzajú schopnosť vystačiť s peniazmi – vzhľadom na úroveň ich aktuálneho príjmu).

Keďže stupnica obsahovala 6 kategórií, bolo nutné uskutočniť transformáciu ordinálnej premennej na binárnu, čomu zodpovedajú tri alternatívne výpočty. Dve z alternatív sú nevhodné na uskutočňovanie záverov o úrovni subjektívnej chudoby v SR, pretože výrazne nadhodnocujú, príp. podhodnocujú zodpovedajúci podiel chudobného obyvateľstva.

Na základe výsledkov sme došli k záveru, že uplatnením logistickej regresie ako nástroja na analýzu subjektívnej chudoby možno za chudobných považovať približne 20% obyvateľstva.

Logitový model je založený na predpoklade logistického rozdelenia rezíduí. K odlišným výsledkom by sme dospeli pri uplatnení napríklad modelov PROBIT (založenom na normálnom rozdelení) alebo GOMBIT (založenom na Gumbelovom rozdelení).

6. Literatúra

[1] ŽELINSKÝ, T., HUDEC, O.: Odhad subjektívnej chudoby na Slovensku založený na distribučnej funkcii rozdelenia príjmov. In: *Forum Statisticum Slovacum*. ISSN 1336-7420.

[2] RAVALLION, M.: *Poverty comparisons: A guide to concepts and methods*. Washington, D. C.: Svetová banka, 1992. ISBN 0-8213-2036-X.

[3] RAVALLION, M.: *Poverty Lines in Theory and Practice*. LSMS Working Paper 133. Washington, D. C.: Svetová banka, 1998. ISBN 0-8213-4226-6.

[4] DUCLOS, J. Y., ARAAR, A.: *Poverty and Equity: Measurement, Policy and Estimation with DAD*. New York: Springer, 2006. ISBN 978-0387-33318-2.

[5] CIPRA, T.: *Finanční ekonometrie*. Praha: Ekopress, 2008. ISBN 978-80-86929-43-9.

[6] ŠÚ SR: *EU SILC 2006, UDB verzia 15/05/2007*.

[7] ŠÚ SR: *EU SILC 2006: Zisťovanie o príjmoch a životných podmienkach domácností v SR*. Bratislava: Štatistický úrad SR, 2007.

Adresa autora:

Tomáš Želinský, Ing.
Ekonomická fakulta
Technická univerzita v Košiciach
Letná 9, 040 01 Košice
tomas.zelinsky@tuke.sk

Príspevok bol napísaný s podporou Vedeckej grantovej agentúry MŠ SR a SAV v rámci riešenia vedecko-výskumného projektu
VEGA 1/0370/08 Regionálny prístup k meraniu sociálneho kapitálu a chudoby.

Obsah

STATISTICA™ 8

analýza dát business intelligence riadenie kvality dáta mining

Nový projekt *STATISTICA* pre školy

Výhody projektu *STATISTICA* pre školy

- ✓ poskytnutie softwaru za zvýhodnených podmienok
- ✓ väčší prístup akademickej obce k profesionálnemu štatistickému softwaru
- ✓ akademická multilicencia
  = ešte výhodnejšia cena = neobmedzený počet inštalácií = software pre všetkých a všade (študenti, zamestnanci, v škole, doma, na notebookoch atď.)
- ✓ software je lokalizovaný do češtiny, samozrejme je i v angličtine a dalších jazykoch
- ✓ možnosť dlhodobejšej a výhodnejšej spolupráce
- ✓ služby pre čo najefektívnějšie užívanie software - kurzy, konzultácie, spracovanie dát.

Projekt už aplikovaly napr. Stredoeurópska vysoká škola v Skalici, Česká zemědělská univerzita, Masarykova univerzita, Mendelova zemědělská a lesnická univerzita, Vysoké učení technické, VŠCHT, Univerzita Karlova - 1. lékařská fakulta a ďalšie.

StatSoft®

StatSoft CR
Podbabská 16, 160 00 Praha 6
www.statsoft.cz
Tel.: (+420) 233 325 006
Fax: (+420) 233 324 005

Stiahnite si
TRIAL VERZIU
www.statsoft.sk/trial

Pokyny pre autorov

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

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

Rozsah: Maximálny rozsah príspevku je 6 strán.

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

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

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

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

Vynechať riadok

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

Vynechať riadok

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

Vynechať riadok

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

Vynechať riadok

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

Vynechať riadok

Vlastný text príspevku v členení:

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

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

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

[1] Písať podľa normy STN ISO 690

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

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

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

FORUM STATISTICUM SLOVACUM

vedecký recenzovaný časopis Slovenskej štatistickej a demografickej spoločnosti

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

Redakcia
Miletičova 3
824 67 Bratislava 24
Slovenská republika

Fax
02/39004009

e-mail
chajdiak@statis.biz
Jan.Luha@statistics.sk

Registráciu vykonalo
Ministerstvo kultúry Slovenskej republiky

Registračné číslo
3416/2005

Evidenčné číslo
EV 2003/08

Tematická skupina
B1

Dátum registrácie
22. 7. 2005

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

ISSN 1336-7420

Ročník
V.

Číslo
1/2009

Cena výtlačku 20 EUR
Ročné predplatné 80 EUR