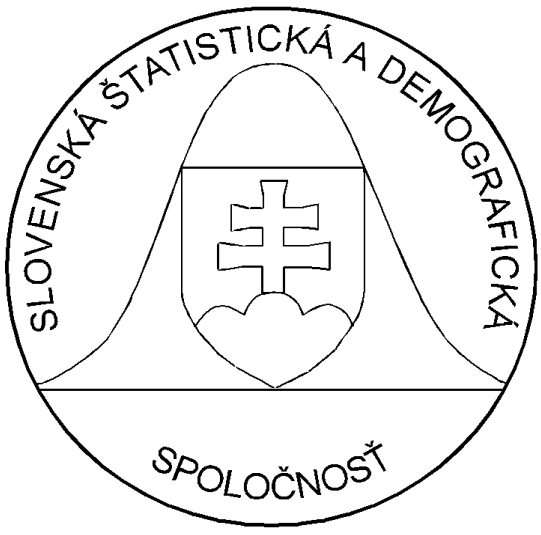
2/2005

FORUM STATISTICUM SLOVACUM



ISSN 1336-7420
52
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)

Konferencia Pohľady na ekonomiku Slovenska 2006,
tematické zameranie: *Makroekonomický vývoj; daň z pridanej hodnoty*
4. 4. 2006, Bratislava

13. Slovenská štatistická konferencia,
tematické zameranie: *Štatistické metódy*
3. – 5. 5. 2006 alebo 20. – 22. 9. 2006 (podľa termínu konania volieb),
Malacky, Hotel ATRIUM

EKOMSTAT 2006 – jubilejná 20. škola štatistiky,
tematické zameranie: *Štatistické metódy v praxi*
21. – 26. 5. 2006, Trenčianske Teplice

PRASTAN 2006,
tematické zameranie: *Pravdepodobnosť, štatistika a numerika*
5. – 9. 6. 2006, Banskobystrický kraj

FERNSTAT 2006,
tematické zameranie: *Aplikovaná, demografická, matematická štatistika, štatistické riadenie kvality*
5. - 6. 10. 2006, Tajov pri Banskej Bystrici

11. Slovenská demografická konferencia,
tematické zameranie: *Migrácia*
3 dni, rok 2007, Košický kraj

Slávnostná konferencia 40 rokov SŠDS,
marec 2008, Bratislava

ÚVOD

Vážené kolegyne, vážení kolegovia,
druhé číslo nového časopisu, ktorý vydáva Slovenská štatistická a demografická spoločnosť (SŠDS) je zostavené z príspevkov, ktoré autori pripravili pre 14. Medzinárodný seminár výpočtová štatistika a na Prehliadku prác mladých štatistikov a demografov. Táto akcia sa uskutočila v dňoch 1. a 2. decembra na Infostate v Bratislave.

Akciu, z poverenia Výboru SŠDS, zorganizoval Organizačný a programový výbor: prof. Ing. Jozef Chajdiak, CSc. – predseda, RNDr. Ján Luha, CSc. – tajomník, RNDr. Peter Mach, Doc. RNDr. Beáta Stehlíková, CSc., Doc. RNDr. Bohdan Linda, CSc., Doc. Dr. Jana Kubanová, CSc., RNDr. Jitka Bartošová, Ing. Vladimír Úradníček PhD., RNDr. Samuel Koróny.

Na príprave a zostavení tohoto čísla participovali: prof. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc.

Recenziu príspevkov zabezpečili: prof. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc., RNDr. Peter Mach, RNDr. Samuel Koróny.

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

Organizátori seminára si považujú za milú povinnosť poďakovať za podporu predsedovi Štatistického úradu SR RNDr. Petrovi Machovi a Infostatu.

Výbor SŠDS

Z HISTÓRIE SEMINÁROV VÝPOČTOVÁ ŠTATISTIKA

Pri príležitosti 14. ročníka semináru Výpočtová štatistika uvádzame srtučnú chronológiu predošlých ročníkov.

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

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

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

Od 5. seminára uskutočneného 5. - 6. 12. 1996 sa už realizuje každoročne ako medzinárodný seminár.
6. medzinárodný seminár Výpočtová štatistika sa uskutočnil 4.- 5. 12. 1997,
7. medzinárodný seminár Výpočtová štatistika sa uskutočnil 3. - 4. 12. 1998,
8. medzinárodný seminár Výpočtová štatistika sa uskutočnil 2. - 3. 12. 1999,
9. medzinárodný seminár Výpočtová štatistika sa uskutočnil 7. – 8. 12. 2000, 10. medzinárodný seminár Výpočtová štatistika uskutočnil 6. – 7. 12. 2001,
11. medzinárodný seminár Výpočtová štatistika sa uskutočnil 5. - 6. 12. 2002
12. medzinárodný seminár Výpočtová štatistika sa uskutočnil 4. - 5. 12. 2003.
13. medzinárodný seminár Výpočtová štatistika sa uskutočnil 2. - 3. 12. 2004.
a 14. medzinárodný seminár Výpočtová štatistika prebieha v dňoch 1. - 2. 12. 2005.

Príspevky 2. seminára boli opublikované v Informáciách SDŠS č. 1/1999 a od 3. seminára sa publikujú v samostatnom Zborníku príspevkov príslušného seminára. Novinkov tohtoročného seminára je publikovanie príspevkov v novozaloženom odbornom časopise SŠDS FORUM STATISTICUM SLOVACUM.

Zameraním seminára je problematika na rozhraní počítačov a štatistiky.
Tematické okruhy posledných seminárov sa nemenia:
- praktické použitie systémov štatistických a príbuzných predmetov,
- práca s rozsiahlymi súbormi údajov,

- vyučovanie výpočtovej štatistiky a príbuzných predmetov,
- praktické aplikácie výpočtovej štatistiky,
- iné.

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

Prípadní záujemcovia z radov mladých štatistikov a demografov (za mladých považujeme štatistikov a demografov pred ukončením vysokej školy) môžu získať informácie na www.ssds.sk , blok akcie a na e-mailových adresách:

chajdiak@statis-ba.sk resp. Jan.Luha@statistics.sk

Semináre z Výpočtovej štatistiky aj Prehliadky prác mladých štatistikov a demografov sa uskutočňujú vo štvrtok a piatok na začiatku decembra. Týždeň pred akciou sa kompletizuje Zborník. Slovensko je malé a nie je odborníkov nazvyš. Prosíme všetkých o aktívne vystúpenie na seminári - za aktívne považujeme hlavne písomné príspevky v Zborníku. Výmena informácií medzi obcou odborníkov SR je jednou z najlacnejších foriem nášho budúceho rozvoja. Informácie o najbližšom seminári získate na www stránke SŠDS http://www.ssds.sk/ resp. http://www.statistics.sk v bloku Slovenská štatistická a demografická spoločnosť.

Prof. Ing. Jozef Chajdiak, CSc.
vedecký tajomník SŠDS
RNDr. Ján Luha, CSc.
člen sekretariátu Výboru SŠDS

Inovácie pohľadom teórií QWERTY a Path Dependency

Vladimír Bačišin

Abstract

As a computer user, you may suffer from pain in your hands due to continual use of the keyboard. This is because we continue to use a badly designed keyboard. Why? And how is this linked to economics? In 1936, August Dvorak designed a keyboard that was easy on the fingers. Yet, the design failed. Why? The decision taken by the people who first used the QWERTY (these are first five keys on the top row) keyboard influenced subsequent users. So, the market was filled with QWERTY users. But what if some users had shifted to a DVORAK keyboard? There would have been a compatibility problem. That is, those used to the QWERTY keyboard would find it difficult to use the DVORAK keyboard. QWERTY, hence, remains the accepted design. This suggests that superior technology need not be the reason for product dominance. Consumer decision process plays a vital role. Economists call this decision process "path dependence". Article descibed QWERTY economis and effects and „path dependence“ concepts in economic theory.

Otázky, ktoré zaujímajú odborníkov, a problémy trápiace profesionálnych vedcov, sú často totožné. Aké otázky historickej vedy sa týkajú problémov profesionálneho publika. Atlantída, tajomstvá slobodomurárov, tajomstvá egyptských pyramíd. Väčšina profesionálnych vedcov sa snažia byť vzdialení od týchto problémov. Jeden z vedeckých problémov najmä v ekonomickej vede – je otázka v závislosti vývoja ekonomických systémov od predchádzajúceho vývoja.

Všetko sa začalo v roku 1985, keď americký historik a ekonóm Paul David zverejnil článok, ktorý sa týkal maličkosti, toho ako sa objavil súčasný štandard klávesníc píšucich zariadení.
V roku 1868 objavili písací stroj. Najprv boli jej klávesy umiestnené v dvoch radoch s postupným zobrazením písmen od A po Z. Avšak prvé modely písacích strojov vyrábané firmou Remington sa vyrábali tak, že ak ste rýchlo stlačili po sebe na klávesnice., ktoré boli vedľa seba, zasekávali sa. Vtedy sa našla verzia klaviatúry, kde najčastejšie po sebe nasledujúce písmená sa nachádzali v rôznych kútoch klávesnice. Tak sa v strede storočia objavila QWERTY klávesnica, ktorá sa stala všeobecnou normou v krajinách s latinskou abecedou.
Paul David poukázal na to, že QWERTY klaviatúra vznikla v dôsledku dočasných a relatívne náhodných technických okolností. Už dva desiatky rokov sa písacie zariadenia tak zdokonaľujú, že zmena klávesníc nie je možná. Avšak QWERTY klávesnica zostala jediným štandardom.
Neskôr americký vynálezca August Dvorak, zapatentoval novú principiálne novú, vedecky zdôvodnenú klávesnicu s iným rozmiestnením písmen abecedy. Práve tu vznikol paradox, ktorého si všimol Paul David. V roku 1940 sa uskutočnili experimenty, ktoré poukázali na to, že klávesnica A. Dvoraka zrýchľuje písanie textov o 20 % až 40 %. Existujú samozrejme názory, že tieto údaje sú nadhodnotené. Nový štandard sa nezačal masovo používať. Klávesnica A. Dvoraka nie je najdokonalejšou. Neskôr boli navrhnuté aj iné rýchlejšie klávesnice. Bez ohľadu na rôzne inovatívne návrhy, nové klávesnice sa objavujú len vďaka niekoľkým „fanatikom“. Väčšina používateľov, alebo „userov“ naďalej používa QWERTY klávesnicu.

Tento jav absolútne odporuje predstave o trhovej konkurencii ako optimálnom mechanizme výberu inovácií. História QWERTY je dôkazom opačného javu. Menej efektívny štandard bol v konkurencii porazený menej efektívnym štandardom.
Americký ekonóm Brian Arthur, ktorý rozvíjal myšlienky Paula Davida, pomenoval tento jav efektom blokácie keď sa uskutočňujú nezvrátiteľné zmeny, pričom len v jednom smere. Keď z množstva konkurujúcich si štandardov víťazí jeden, návrat k množine štandardov nie je prakticky možný. Znamená to, že je nevyhnutné víťazstvo jedného zo štandardov, ale neexistuje objektívna zákonitosť, ktorý zo štandardov zvíťazí. Veľkú úlohu tu hrá historická náhoda, ktorá kdesi na začiatku skúmaného procesu určuje všetky nasledujúce udalosti.
V situácii neurčitosti majú veľmi veľký význam očakávania ľudí. Špecifický systém mohol zvíťaziť nad konkurentmi jednoducho preto, lebo , že kupujúci očakávali toto víťazstvo, poznamenával P. David. Rast očakávaní môže byť dôsledkom veľmi, zdalo by sa bezvýznamných udalostí.
V histórii štandardu QWERTY zohral veľkú úlohu marketingový trik, ktorý použil výrobca písacích strojov Remington. V americkom meste Cincinatti sa 25. júla 1888 uskutočnila súťaž dvoch pisárov – jeden písal na Remingtone a ďalší na kaligrafe. Víťazom v súťaži sa stal Remington s QWERTY klávesnicou, ktorý sa stal ihneď veľmi populárny, ktorý sformoval očakávania, podľa ktorých práve to písacie zariadenie je najlepšie z najlepších.
História víťazstva QWERTY klávesníc nad efektívnejšími štandardmi sa môže stať niečím málo významným. Avšak skúmanie histórie technických štandardov, ktoré sa začalo po pionierskych prácach Paula Davida a Briana Arthura, ukázalo, že QWERTY efekty sa stretávajú doslova na každom kroku.
Pod QWERTY efektmi sa v súčasne vedeckej literatúre chápu všetky druhy relatívne neefektívnych, ale stabilne sa zachovávajúcich sa štandardov. Tieto efekty sa dajú objaviť dvomi cestami. Buď sa porovnávajú dva typy súčasne reálne existujúcich štandardov, alebo sa porovnávajú realizované technické inovácie s potenciálne možnými, ale neralizovanými inováciami. Hoci sa súčasná ekonomika globalizuje a unifikuje, ale vo svete sa zachovávajú rôzne technické štandardy, ktoré do seba nezapadajú. Jedným z klasických príkladov je doprava po ľavej strane cesty vo väčšine krajín a doprava na pravej strane vo Veľkej Británii. To výrobcov automobilov núti, aby na jedných automobiloch montovali volanty sprava a na druhých volanty zľava. Povedzme si o menej známom, ale dôležitom rozdiele v šírke železničných koľají.
Šírka železničných koľají je v rôznych krajinách rôzna. Základnými sú tri štandardy Prvý je 1067 mm - známy ako Kapská koľaj (Japonsko, JAR, Sachalin a ďalšie krajiny tretieho sveta), 1435 mm - Európska (Stephensonova) koľaj (viac ako polovica železníc sveta), 1520 – 124 mm – Ruská koľaj (väčšina postsovietskych krajín). Tento rozdiel vidieť najmä vtedy ak cestujete vlakom a prechádzate východné hranice. Tam sa vlaky zdržiavajú kvôli výmene kolies. Je európska železnica najefektívnejšia? Vôbec nie. Technici považujú za vhodnejšiu širokorozchodnú železnicu.
Európsky štandard sa náhodne objavil v začiatkoch rozvoja železničnej dopravy. Keď v roku 1825 známy anglický inžinier George Stephenson vynašiel lokomotívu, svoju normu jednoducho vzal z anglických konských povozov. Medzi ne patrila dĺžka osi, teda je vzdialenosť medzi kolesami 4 stopy a 8,5 palca. Stephenson využil túto normu v roku 1830 pri výstavbe prvej železničnej trate Liverpool – Manchester. Tá sa stala modelom pre výstavbu železníc nielen vo Veľkej Británii, ale aj celej kontinentálnej Európy a Severnej Ameriky.
Je však známe, že konkurenti Stephensona ponúkali to, aby sa vzor prijala norma s rozchodom koľajníc 5 stôp a 6 palcov (1676 mm). Ak by objednávku na výstavbu železnice vyhrali konkurenti Stephensona, norma rozchodu železničných koľajníc by zostala úplne iná.

Rusko začalo stavať železnice v čase keď panoval cár Mikuláš I. Rusko si vybralo rozchod koľajníc 5 stôp. Mali lepšiu kapacitu, ale znamenali horšiu prepravu cez európske koľajnice. Zabezpečovala väčšiu priepustnosť, ale komplikovala aj prekonanie európskej hranice. Oba tieto faktory hrali pozitívnu úlohu pre pripravenosť na vojnu s európskymi krajinami v minulosti, ale v súčasnosti sú prekážkami. Že by QWERTY efekty vznikli len v ranných štádiách vývoja ekonomiky? Nie. Prejavujú sa aj v neskorších etapách vývoja vedy a techniky. Ako príklad môže slúžiť 500 riadkový štandard pre televíziu v USA a 800 riadková norma v Európe. Ďalším príkladom je víťazstvo štandardu VHS nad BETA.

V porovnaní s výskumom súťaže rôznych štandardov oveľa rozumnejšou je analýza neuskutočnenej ekonomickej histórie. Myslíme tým tie inovácie, ktoré v dôsledku momentálnej konjunktúry zahatali cestu k životu efektívnejším inováciám. Myšlienka porovnávania reálne uskutočnených a potenciálne možných technologických stratégií bola po prvý raz vyslovená v roku 1964 v škandálne známej knižne vydanej dizertačnej práce z roku neskoršieho nositeľa Nobelovej ceny za ekonómiu z roku 1993 Roberta Williama Fogela Railroads and American economic growth: essays in econometric history (Železnice a ekonomický rast Ameriky: eseje z ekonometrickej histórie). Rodičia Roberta Williama Vogela sa do USA prisťahovali v roku 1922 z ukrajinského mesta Odesa do USA.

Robert William Fogel vypočítal čo by sa stalo ak by sa v USA namiesto železníc stavali riečne kanály a rozvíjala by sa doprava konskými záprahmi. Ukázalo sa, že živé kone nie sú oveľa lepšie ako železné a výstavba železníc urýchlila vývoj USA len o jeden rok. Okolo knihy Roberta Williama Fogela sa rozpútala hlasná diskusia. Kritici poukazovali na to, že jeho výpočty sú len relatívne, pretože ťažko sa dá zmerať niečo čo nebolo. Po búrlivej diskusii sa Robert William Fogel prestal zaoberať touto témou a presmeroval svoje vedecké záujmy na ekonomiku otroctva a takmer sa prestal zaoberať „alternatívnou históriou“. Jeho skúsenosti neprepadli, ale začali sa nim zaoberať QWERTY- ekonómovia.

Hoci sa nasledovníci Paula Davida nesnažia hodnotiť alternatívne technologické stratégie, ale veľmi používajú kvalitatívne porovnanie reálneho s potenciálne možným. Robert William Fogel dokázalo, že v histórii zvíťazil efektívnejší prístup, ale QWERTY ekonómovia poukazujú na to, že niekedy víťazia aj neefektívne prístupy.

Veľmi zaujímavým dôvodom pre QWERTY prognózy vytvorila atómová energetika. Súčasné mierové využitie atómovej energie je vedľajším efektom produktom studenej vojny. Prvé elektrárne postavené v rokoch 1950 – 1960 mali za úlohu predovšetkým ukázať, že atómová energetika má aj mierové využitie. V dôsledku toho sa v americkej energetike využívali reaktory, ktorých fungovanie je založené na takzvanej ľahkej vode, ktoré sú vlastne adaptovanými reaktormi z atómových ponoriek. Existuje názor, že reaktory postavené tak, aby boli ochladzované plynom by boli efektívnejšie.

Po mnohých výskumoch QWERTY efektov, mnohí ekonómovia historici prišli k záveru, že mnohé symboly technického progresu získali súčasne dobre známu podobu v dôsledku v mnohom náhodných okolností a že žijeme v mnohom nie lepších svetov.

Najdôležitejšia z myšlienok, ktoré obohatili prvotnú koncepciu Paula Davida, pozostáva v tom, že víťazstvo primárne vybraných štandardov / noriem nad všetkými inými – dokonca aj nad efektívnejšími sa dá pozorovať len v histórii vývoja technológií. Súčasní historici ekonómie predpokladajú, že súčasný vývoj spoločnosti nie je závislý len od vývoja pracovných nástrojov, ale aj od objavenia sa nových inštitúcií ako sú ekonomické organizácie, právne a etické normy. Ukazuje sa, že aj medzi týmito pravidlami nie vždy víťazia tie najefektívnejšie. O stabilnom zachovaní neefektívnych technológia noriem sa začalo hovoriť ako o prejave závislosti od predchádzajúceho vývoja Path Dependency.

Fakticky, akýkoľvek vzor technologických QWERTY efektov má povinnú inštitucionálnu podporu. Konkurujú si nie technológie, ale organizácie, ktoré ich prinášajú Napríklad

víťazstvo úzkeho rozchodu koľajníc Stephensona nie na efektívnejším štandardom širokej koľaje, je víťazstvo nemej efektívne firmy, ktorá mala väčšie šťastie.
Veľmi zaujímavý príklad závislosti od predchádzajúceho rozvoja poukázal v deväťdesiatych rokoch americký ekonóm Raphael La Porta, ktorý pripravil s kolegami komparácie právnych systémov v rôznych krajinách. Ako je známe, takmer všetky národné právne systémy sa dajú rozdeliť na dve triedy občianske (rímsko-germánske) právo pochádzajúce priamo z rímskeho práva a všeobecné (anglosaské) právo odrážajúce „amatérsku“ činnosť stredovekých Angličanov. P. La Porta a jeho kolegovia presvedčivo dokázali, že tradície všeobecného práva sú lepšie pre oblasť vlastníckych práv, napríklad v ochrane malých akcionárov firiem pred zneužívaním postavenia manažérov. Avšak tieto výhody nevedú k tomu, aby väčšina krajín s kontinentálnym právom prešla k anglosaskému modelu práva.
Teória závislosti od predchádzajúceho vývoja a blízke k nej vedecké výskumy z alternatívnej histórie sú založené na metavedeckej paradigme synergetiky. Reči ide o myšlienkach známeho belgického chemika Ilyu Prigozhina, ktorý vytvoril teóriu samoorganizácie poriadku z chaosu. V súlade s ním rozpracovaným princípom, rozvoj spoločnosti nie je presne predurčený. , teda nefunguje podľa zásady „ináč to nebude“. V skutočnosti sa pozoruje striedanie období evolúcie, keď sa vektor rozvoja nedá zmeniť (pohyb podľa atraktora) a bodov javu označovaného v angličtine point of bifurcation, prekladaného ako bod rozvetvenia, alebo rozdvojenia, v ktorom existuje možnosť výberu.
Keď QWERTY ekonómovia hovoria o historickej náhode prvotného výberu, oni posudzujú práve historické body rozdvojenia, alebo rozvetvenia, teda tie momenty keď dochádza určitej jednej možnosti z vejára rôznych alternatív. Výber v týchto situáciách sa prakticky vždy odohráva v situáciách neurčitosti a nestability bilancie sociálnych síl. V bode rozvetvenia sa môžu ukázať osudové dokonca úplne malé okolnosti – podľa princípu „motýľa Bradbury.“
Súčasní autori, ktorí sa venujú závislosti od predchádzajúceho vývoja, sa podčiarkuje množstvo faktorov, ktoré vyvoláva závislosť. Paul David Poukazoval, že v závislosti od predchádzajúceho vývoja v ekonomike existuje technická vzájomná závislosť. Systémy zariadení a návyky pracovníkov vytvárajú jednotný systém, ktorá pozostáva zo vzájomne doplňujúcich sa prvkov. Ukazuje sa, že je veľmi problematické kvalitatívne zmeniť čo len jeden z jeho prvkov. Napríklad hoci železničné trate sa neskladajú čo len z jedného detailu z čias D. Stephensona, šablóny pre ich výrobu koľajníc sú tie isté.
Ďalším faktorom je rast ziskov v závislosti od veľkosti. Použitie akéhokoľvek štandardu je o to výhodnejšie, o čo častejšie sa používa. Pri výstavbe nových železničných tratí, využitie starého štandardu uľahčuje zapojenie nových tratí s starým trasám. Nové železničné trate preto vždy využívajú tie isté štandardy. Je to aj vtedy keď si inžinieri a technici uvedomujú zastaranosť štandardov.
Analogický jav sa dá použiť k inštitúciám. Zvykne sa pomenovať sieťovými efektmi. Napríklad každá krajina by mohla vytvárať svoj vlastný neopakovateľný systém práva, ale komplikovalo by to medzinárodnú deľbu práce. Preto rôzne krajiny unifikujú normy regulujúce napríklad obchod. Deje sa to v rámci Svetovej obchodnej organizácie (WTO). Dokonca, ak sú aj tieto normy neoptimálne, a sú rozšírené v malom množstve krajín, môže sa stať, že vytlačia efektívnejšie normy, ktoré sú rozšírené v menšom počte krajín.
Ďalším faktorom potvrdzujúci závislosť od predchádzajúceho vývoja je opotrebovanie investičného majetku, kvázinenávratnosť investícií. Celá záležitosť pozostáva v tom, že morálne zostarnuté zariadenia môžu fungovať ďalej, pretože vysoké investície do nich sa ešte nevrátili. Aj preto sa dá v niektorých podnikoch naraziť na stroje, ktoré fungujú od čias prvej Československej republiky. Trvanie tohto faktora je obmedzené dobou fyzického opotrebovania morálne zastaralého kapitálu, ktorý sa bude musieť raz predsa len odpísať.
Tento faktor pripúšťa netriviálnu inštitucionálnu interpretáciu. Sociálne normy (napríklad etické hodnoty) môžu tiež morálne zastarať. Vymeniť je ich oveľa ťažšie, ako investičný

majetok: Formovanie osobnosti sa končí v mladosti, ale neskôr adaptácia môže len veľmi málo zmeniť v normách získaných v mladosti. Dĺžka fungovania sociálneho kapitálu, teda nemennosť prvotnej socializácie je dlhšia ako fungovanie investičného majetku. Priemerný život človeka je dlhší ako život stroja.

Posledný fakt je spojený s nerovnomernosťou rastu zisku. Keď si konkurujú rôzne technológie / inštitúcie, primárny rast hraničnej úžitkovosti môže ísť jednou cestou a potom druhou.

Veľká pozornosť ekonómov koncepcii závislosti od predchádzajúcej cesty rozvoja je spojená so zlyhaniami trhu. Ak je víťazstvo horších noriem nad lepšími spojené s nepredvídanosťou trhu, je to veľkým argumentom pre štátnu reguláciu ekonomiky. Zástancovia trhu však navrhujú inú interpretáciu. Podľa nich sú zlyhania trhu spojené so zlyhaniami štátnej regulácie. Pravdu majú obe strany. Zlyhanie klaviatúry je zlyhaním trhu. Naopak zlyhanie atómového reaktora je zlyhaním vlády.

David, Paul A, 1985, "Clio and the Economics of QWERTY," American Economic Review, American Economic Association, vol. 75(2), pages 332-37.

David, Paul A., 1994. "Why are institutions the 'carriers of history'?: Path dependence and the evolution of conventions, organizations and institutions," Structural Change and Economic Dynamics, Elsevier, vol. 5(2), pages 205-220.

Paul A. David, 2004. "Understanding the emergence of 'open science' institutions: functionalist economics in historical context," Industrial and Corporate Change, Oxford University Press, vol. 13(4), pages 571-589.

David P.A. Understanding the Economics of QWERTY: The Necessity of History // Economic History and the Modern Economist. Ed. by William N. Parker. New York: Basil Blackwell, 1986.

Mgr. Vladimír Bačišin
(Autor je doktorandom Prognostického ústavu SAV)
Bačišin / Tkáč, s. r. o.
Hviezdoslavovo nám. 17
811 02 Bratislava
www.bacisintkac.sk

VALIDITY OF THE LOGARITHM-NORMAL MODEL OF HOUSEHOLD INCOME DISTRIBUTION IN THE CZECH REPUBLIC

Jitka Bartošová

Abstract. Statistical models of income distribution provide an evaluation of the living standards of the population of the whole country as well as comparison of the living standards of different social classes or regions. In the Czech Republic, the most frequently used type of theoretical model, the logarithm-normal distribution – especially its three-parameter version, used to be suitable for approximation of household income distribution both, on whole and of individual social classes. After revolution, the transformation to the market economic system, mainly the formation of new income sources and the process of significant differentiation of wages, have caused a significant changes in the income distribution. This paper concentrates on verification of validity of the statistical model of income distribution presently used in the Czech Republic, and detection of income values that cause the contamination of the model.

Keywords. Estimate of parameter, income distribution, logarithmic-normal model, validity of parametric model.

1 Parametric model of income distribution after Velvet Revolution

To characterize empirical households' income distribution, it's often convenient to use a parametric model. One of the most used parametric models of households' income distribution is the logarithm-normal distribution. This model is also used in economics to illustrate distribution of wages and job standardization. The logarithm-normal distribution competes with Weibull's distribution in the field of stipulation of the time to failure of the device. Besides, it's also frequently used for quality control and in the theory of reliability.

Initially, development of the logarithm-normal distribution was studied by AITCHISON and BROWN. They were interested mainly in its application in astronomy, biology, sociology, economics or simulation of physical processes. These authors also formulated special reasons concerning the problem of the logarithm-normal distribution and income models. The biggest "competitor" of the logarithm-normal distribution in income modeling is Pater's distribution. The logarithm-normal curve corresponds to the empirical income distribution in a large central area, while in extremes it significantly diverges. On the contrary, Paret's curve is a suitable model of income distribution in extreme values (see JOHNSON, KOTZ and BALAKRISHNAN). It means that it's suitable to use the logarithm-normal model for representation of income distribution of the majority of households of the central part of range and Paret's distribution for extremes.

As a theoretical model of income distribution, the logarithm-normal distribution has been used so far. Recently used type of theoretical model was derived from the character of a particular feature and from a longtime experience with its behavior before revolution. Especially the three-parameter logarithm-normal distribution has represented a good approximation of income distribution for most of social classes. Since the three-parameter logarithm-normal distribution with parameters μ, σ^2 and γ, where γ is the theoretical

minimum, represented a good approximation of income distribution in the Czech Republic before revolution, it's necessary to verify validity of the model after the revolution.

Before revolution, planned economy in the Czech Republic experienced high homogeneity of income in the population in all social classes. In presence, the transformation to market economic system has caused a significant change in income distribution. The variety of income sources and present process of differentiation of wages brings about discrepancies between empirical income distribution and the theoretical model. There are values of income that may be concerned as outliers and that causes contamination of the model (see ANTOCH and VORLÍČKOVÁ, JUREČKOVÁ, BARTOŠOVÁ, 9). These differences are still deepening and in a different rate and inertia they are progressively reflected in both income distribution of the whole population without regarding social classes and income distribution of some social classes (see BARTOŠOVÁ, 5-8).

There is some inertia in income distribution, so its changes would noticeably display in a term of several years after revolution. The first sample survey focusing on distribution of income of population, Mikrocensus, in which we may anticipate significant changes in its results, was carried out by Czech Statistical Office in 1996.

2 Point estimate of parameters of the logarithm-normal model

While making a statistical model it is important both, to find out a theoretical distribution function that would characterize empirical frequency distribution and to choose suitable methods to calculate parameters of the model. An important of the construction of the model is a choice of suitable parameters (see ANDĚL, 2-3). As to assumed changes in income distribution, such as high variability, contamination etc., that came after Velvet Revolution, it's necessary to choose a method of parameters estimates that gives good results also under these new conditions. Such a method can be the method of maximal likelihood. Maximal likelihood estimate of the parameters μ, σ^2, γ of the three-parametric logarithm-normal model is for sufficiently smooth likelihood function $L\left(\mu, \sigma^2, \gamma | x\right)$ based on the solution of the likelihood equation array

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \ln\left(x_i - \hat{\gamma}\right) \;\wedge\; \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \left[\ln\left(x_i - \hat{\gamma}\right) - \hat{\mu}\right]^2 \;\wedge\; 0 = \sum_{i=1}^{n} \frac{\ln\left(x_i - \hat{\gamma}\right) - \hat{\mu} + 1}{x_i - \hat{\gamma}}. \qquad (1)$$

The maximal likelihood estimate of parameter γ of three parameter logarithmic-normal distribution can be numerically calculated as the maximum of the modified logarithmic likelihood function. In the case of sample size n, we have

$$\ell(\gamma) = -n\left[\hat{\mu}(\gamma) + \frac{1}{2}\ln\hat{\sigma}^2(\gamma)\right], \qquad (2)$$

where $\hat{\mu}(\gamma)$ and $\hat{\sigma}^2(\gamma)$ are maximal likelihood estimates of parameters and γ is a chosen value of theoretical minimum of the model.

Considering the character of the particular feature, I obtained the estimate by finding the maximum value of the function $\ell(\gamma)$ in the interval $\left(-x_{\max}, x_{\min}\right)$, where $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of incomes. The task was solved by the iteration

method – on a lattice, which density was increased in each iteration step. Corresponding iteration procedure was implemented in programme MATLAB.

Because it's not possible to determine maximal likelihood estimate of the theoretical minimum directly and it's needy to use numerical maximization, I estimated this parameters by some more methods. To determine the theoretical minimum I used following null (two-parametric model), sample minimum, Cohen's method, method of maximal likelihood and method of likelihood ratio minimization. First three estimate methods are very simple, but at the same time they are totally not sensitive or only a little sensitive to character of empirical income distribution. Thus it's interesting to compare their results with results of other two iteration methods.

To estimate parameter γ by numerical minimization of likelihood ratio $LR(\gamma)$ we can use the same iteration procedure, that was used to numerically maximize function $\ell(\gamma)$,

$$LR(\gamma) = 2\{\ell[\bar{\pi}(\bar{\mu}(\gamma), \bar{\sigma}^2(\gamma)] - \ell(\vec{p})\}. \tag{3}$$

Maximal likelihood estimates of parameters μ, σ^2 and γ of the three-parameter logarithm-normal model of net annual monetary incomes per household are in the table 1, estimates of the parameters of the models of incomes per one member of the household is

Table 1: Maximum likelihood estimates of parameters μ, σ^2 and γ
(Income per household)

Social class	$\bar{\mu}$	$\bar{\sigma}^2$	$\bar{\gamma}$	$-\ell(\bar{\gamma})$
Worker	12,1841	0,12416	-30125	98665
Self-employed	12,2498	0,33054	-1256	20445
Employees	12,2060	0,22624	-10753	79266
Self-employed farmers	12,0965	0,46730	6066	1535
Farmers–member of cooperative	12,1635	0,14463	-14197	2183
Retired with EA members	11,7931	0,18621	34829	12661
Retired without EA members	11,0184	0,20975	11972	88564
Unemployed	11,3404	0,32690	-8185	2803
Others	11,0142	0,49000	3272	2515
All	11,7721	0,39144	2666	318161

Table 2: Maximum likelihood estimates of parameters μ, σ^2 and γ
(Income per head)

Social class	$\bar{\mu}$	$\bar{\sigma}^2$	$\bar{\gamma}$	$-\ell(\bar{\gamma})$
Worker	10,8854	0,18693	4231	88975
Self-employed	11,0041	0,48396	7596	18601
Employees	11,0311	0,30152	9740	72135
Self-employed farmers	10,6747	0,84755	10535	1388
Farmers–member of cooperative	11,0236	0,11049	-4379	1965
Retired with EA members	11,0320	0,09850	-391	11413
Retired without EA members	10,7734	0,05378	4897	80557
Unemployed	10,2194	0,30573	1959	2503
Others	9,9224	0,65831	6653	2292
All	10,9083	0,21206	4555	285220

in the table 2. In the tables, there's also the value of the reached maximal of the logarithmic likelihood function $\ell(\bar{\gamma})$.

3 Validity of the logarithm-normal model

Quality of the logarithm-normal model of income distribution in 1996 is determined by quality of estimate of its parameters. Essentially the best estimate is the maximal likelihood estimate of all three parameters of the model. All the other models, which resulted from combining maximal likelihood estimates of μ and σ^2 with different estimate of γ, are compared by way of likelihood ratio, i.e. statistic

$$LR(\mu,\sigma^2,\gamma|n) = 2[\ell(\vec{p}|n) - \ell(\vec{\pi}(\mu,\sigma^2,\gamma)|n)], \tag{4}$$

where $\vec{p}$ is the vector of income empirical probability, $\vec{\pi}(\mu,\sigma^2,\gamma)$ is the vector of probabilities of occupation of particular classes and $\ell(\vec{p}|n)$, $\ell(\vec{\pi}(\mu,\sigma^2,\gamma)|n)$ are corresponding logarithm-normal likelihood functions. Results are in Tables 3 and 4. Besides information about likelihood ratio, there are values of corresponding quantiles χ^2.

Table 3: Comparison of agreement of empirical distribution with logarithm-normal models (incomes of whole households)

Social class	$\gamma = 0$	$\gamma = x_{min}$	γ–Cohen	γ – LR	χ^2
Worker	475,9	907,3	300,8	232,8	108,6
Self-employed	77,4	108,6	81,3	77,1	59,3
Employees	145,2	264,6	114,1	109,4	99,6
Self-employed farmers	31,3	34,3	32,0	31,0	22,4
Farmers–member of cooperative	14,8	28,8	15,0	14,8	26,3
Retired with EA members	32,6	38,8	29,6	21,4	51,0
Retired without EA members	3874,5	3690,6	4027,7	3698,5	108,0
Unemployed	29,3	51,7	26,7	23,0	28,9
Others	26,9	21,8	33,3	25,3	27,6
All	3239,8	3233,1	3311,7	3224,4	167,5

Table 4: Comparison of agreement of empirical distribution with logarithm-normal models (incomes per a member of household)

Social class	$\gamma = 0$	$\gamma = x_{min}$	γ–Cohen	γ – LR	χ^2
Worker	170,6	279,7	174,7	163,1	108,6
Self-employed	94,6	55,3	97,2	55,8	59,3
Employees	167,4	189,1	104,5	104,2	99,6
Self-employed farmers	29,0	23,0	23,8	22,7	22,4
Farmers–member of cooperative	11,5	41,8	12,4	11,0	26,3
Retired with EA members	86,0	120,3	101,7	84,8	51,0
Retired without EA members	1520,6	1672,2	2675,8	1519,3	108,0
Unemployed	17,8	25,1	18,0	17,8	28,9
Others	76,9	38,3	58,4	44,4	27,6
All	2702,5	2533,3	3253,9	2534,2	167,5

2 Conclusions

As we can see in Tables 3 and 4, in almost all social classes works $LR \approx \chi^2$, so that our logarithm-normal curves are mostly good approximations of empirical distribution of household' incomes in 1996. Only in the case of retired without economically active

members and the case of all households (without regarding social classes) the logarithm-normal model is totally unsuitable as a model for empirical distribution. In both cases above statistics LR was more then ten times higher than corresponding quintile. Moreover graphs of kernel estimates of income distribution density per the whole household are two-peaked in both cases and it clearly signalizes commixture of two one-peaked curves. So that it's not possible to find such one-peaked parametric model that would well approximate this empirical distribution. On the other hand both corresponding income distributions per a member of household are one-peaked, therefore one should search for other, better parametric models.

References

1. AITCHISON, J., BROWN, J.A.C.: The Lognormal Distribution with Special Reference to Its Uses in Economics, Cabridge University Press, Cambridge, 1957.
2. ANDĚL, J.: Statistické metody, Matfyzpres, Praha, 1993.
3. ANDĚL, J.: Základy matematické statistiky, Univerzita Karlova v Praze, Praha, 2002.
4. ANTOCH, J., VORLÍČKOVÁ, D.: Vybrané metody statistické analýzy dat, Academia, Praha, 1992.
5. BARTOŠOVÁ, J.: Příjmové modely. In J. Chajdiak, J. Luha and S. Koróny, eds., 12. Medzinárodný seminár Výpočtová štatistika, Zborník príspevkov, Bratislava 5.-6. 12. 2003, s. 7-11, SŠDS, Bratislava, 2003.
6. BARTOŠOVÁ, J.: Statistic Model of Households' Annual Income Distribution in the Czech Republic. In J. Antoch, editor, Compstat 2004, 16th Symposium of IASC, Praque, August 23-17, 2004, Books of Posters and Abstracts [CD-ROM], s. 1-8, Czech Statistical Society, for IASC, Praha, 2004.
7. BARTOŠOVÁ, J.: Příspěvek k analýze rozdělení příjmů domácností v ČR. In J. Antoch and G. Dohnal, editors, Robust 2004, Sborník prací 13. letní školy JČMF Robust 2004 ve dnech 7.-11. června v Třešti, s. 451-458, JČMF, Praha, 2004.
8. BARTOŠOVÁ, J.: Statistický model příjmových rozdělení. Acta Universitatis Bohemiae Meridionales. **7**(2): s. 39-46, Jihočeská univerzita, České Budějovice, 2004.
9. BARTOŠOVÁ, J.: Contamination Level Estimate of Household Income Distribution by Distant Observations in the Czech Republic. Forum Metricum Slovakum **8**, s. 74-78, SŠDS, Bratislava, 2004.
10. JOHNSON, J.N.L., KOTZ, S., BALAKRISHNAN, N.: Continuous Univariate Distributions, vol. **1**, 2nd edition, J. Wiley, New York, 1994.
11. JUREČKOVÁ, J.: Robustní statistické metody, Karolinum, Praha, 2001.

Contact address

RNDr. Jitka Bartošová, katedra managementu informací, Fakulta managementu VŠE Praha, Jarošovská 1117/II, 377 01 Jindřichův Hradec, Czech Republic, tel: +420 384 417 221, e-mail: bartosov@fm.vse.cz

Monte Carlo simulačná štúdia pre metódu ANOVA v systéme SAS®

Mária Bohdalová

Abstract

Analysis of variance (ANOVA) is a statistical technique widely used in a variety of disciplines, including, but no limited to, finance research, marketing, psychology, etc. Like many other parametric statistic techniques, one fundamental assumption for ANOVA is that the dependent variable is normally distributed. Another important assumption for ANOVA is that the groups come from populations with equal variances. But how serious are the consequences if the assumption about data normality is violated, or if the assumption of equal population variances is violated? In this paper we will presented a SAS Monte Carlo simulation study for assessing the consequences of data non-normality and unequal population variances on the Type I error rate of ANOVA analysis.

Úvod

Kvantitatívne modely matematické alebo štatistické charakterizuje determinovanosť. Ide hlavne o tie modely, pomocou ktorých hľadáme najlepšie, resp. optimálne riešenie. V manažérskej praxi sa však často stretávame so situáciami, v ktorých je hľadanie optimálneho riešenia nereálne, napr. nie sú splnené základné podmienky používaných metód. Na riešenie takýchto situácií sa s obľubou využívajú simulačné modely. Simulovanie (napodobňovanie) reálnych problémov patrí k často používaným prístupom, ktorý uľahčuje rozhodovanie manažérov. Od klasických modelov sa simulačné modely líšia tým, že sú zamerané na popis skúmaného problému a nie na nájdenie jeho optimálneho riešenia.

Simulácie môžeme charakterizovať ako numerické techniky určené pre riadenie experimentov na počítači. Metóda Monte Carlo je známa simulačná metóda, ktorá používa veľký počet náhodne generovaných vzoriek z daného pravdepodobnostného rozdelenia, ktoré slúžia na počítačovú simuláciu riešenia rôznych manažérskych problémov z oblasti financovania, projektovania, predaja, personalistiky, psychológie, matematiky, fyziky [Fan02] a pod.

V štatistickej teórii sa stretávame s dvoma základnými typmi metód: parametrickými a neparametrickými. Parametrické štatistické metódy sa vyznačujú tým, že ich platnosť je podmienená splnením istých teoretických predpokladov. Keď spracúvané údaje spĺňajú teoretické predpoklady použitých metód, štatistické metódy ponúkajú platné a účinné odhady štatistík ich pravdepodobnostného rozdelenia. Na druhej strane, keď teoretické predpoklady aktuálne skúmané údaje nespĺňajú, potom platnosť spoľahlivých odhadov štatistík ich pravdepodobnostného rozdelenia je často diskutabilná a neistá. V dôsledku toho buď strácame dôveru k teoretickým odhadom, alebo ak sa naslepo spoliehame na teóriu, môžu byť naše závery nesprávne, pretože isté veľmi dôležité predpoklady konkrétne použitej metódy neboli splnené. V takýchto analytických situáciách ponúka simulačná metóda Monte Carlo veľmi užitočné závery pre kvantitatívny výskum. Simulačná metóda Monte Carlo uprednostňuje empirické odhady štatistík pravdepodobnostného rozdelenia vzorky, pred teoretickými očakávaniami o týchto hodnotách. Podstatou simulačnej metódy Monte Carlo je, že generuje početné náhodné scenáre pre skúmanú vzorku. Takto získané výsledky sa asymptoticky blížia k teoretickým hodnotám. Dá sa preto povedať, že pomocou simulačnej metódy Monte Carlo možno demonštrovať, ako sa priblížiť k teoretickým výsledkom. Simulačnú metódu Monte Carlo je vhodné použiť ako alternatívnu metódu aj vtedy, keď štatistické teórie nie sú dostatočné, alebo keď jednoducho neexistuje štatistická teória pre riešenie daného problému. V takýchto prípadoch simulačná metóda Monte Carlo môže byť jedným z realizovateľných postupov poskytujúcich odpoveď na dané problémy kvantitatívneho výskumu. V tomto príspevku sa zameriame na situáciu, keď nie sú splnené predpoklady štatistickej metódy.

Samotnú Monte Carlo simulačnú štúdiu vykonávame v nasledujúcich krokoch:

1. Navrhneme otázky, na ktoré chceme simulačnou štúdiou získať odpovede (určenie cieľa simulačnej štúdie)

2. Navrhneme takú Monte Carlo simulačnú štúdiu, ktorá nám poskytne želané odpovede.
3. Náhodne vygenerujeme údaje na základe konkrétnych štatistík.
4. Implementujeme kvantitatívne (štatistické) metódy.
5. Vyčíslime a zhromaždíme štatistiky, ktoré nás zaujímajú v každej realizácii simulačnej štúdie.
6. Opakujeme kroky (3)-(5) vhodný počet krát (napr. 1000 – 10 000 krát)
7. Analyzujeme vyčíslené a zhromaždené štatistiky.
8. Vyhodnotíme empiricky získané výsledky v Monte Carlo simulačnej štúdii.

Simulačná metóda Monte Carlo je výpočtovo náročná metóda, ktorá vyžaduje nielen znalosť programovania, štatistických a matematických metód ale i systém, ktorý poskytuje kombináciu zabudovaných štatistických/matematických procedúr a má možnosť programovania. Napríklad jednou možnosťou je použiť programovací jazyk FORTRAN. Tento jazyk nemá zabudované štatistické/matematické procedúry a je ich potrebné buď naprogramovať, alebo treba pristupovať do knižnice naprogramovaných štatistických procedúr, čo môže spôsobiť komplikácie. SAS® System je v porovnaní s programovacím jazykom FORTRAN výhodnejší, pretože ponúka v rámci jedného systému veľký výber štatistických procedúr v SAS/STAT a SAS/ETS softvéri [Sta00], matematické funkcie a univerzálne programovacie možnosti ponúka v SAS/BASE, SAS Macro Facility a v SAS/IML softvéri. Táto kombinácia možností robí SAS® System vhodným nástrojom pre riadenie Monte Carlo simulácií [Fan02]. Pre účely tohto príspevku som použila softvérový systém SAS® verziu 9.1.

Monte Carlo simulačná štúdia pre vyhodnotenie efektov nesplnených predpokladov pre metódu ANOVA

Uvažujme veľmi často používanú a obľúbenú metódu analýzy rozptylu (ANOVA). Táto metóda umožňuje porovnávať stredné hodnoty dvoch alebo viacerých súborov [Pac03]. Jej použitím môžeme dostať odpoveď na otázku, či na zvolenej hladine významnosti nezamietneme alebo zamietneme nulovú hypotézu :

$$H_0\colon\ \mu_1=\mu_2=\ldots=\mu_k \quad \text{pre } k\geq 2\,, \tag{1}$$

Pričom alternatívna hypotéza znie:

$$H_1\colon \text{aspoň dve stredné hodnoty sa nerovnajú}\,. \tag{2}$$

Základné súbory (skupiny), ktorých stredné hodnoty porovnávame, v praxi často vzniknú rozdelením jedného základného súboru podľa k úrovní určitého faktora, pričom $k\geq 2$. Metódu analýzy rozptylu môžeme v praxi použiť ak sú splnené tri základné podmienky [Pac03]

1. výbery sú nezávislé,
2. vo všetkých porovnávaných k základných súboroch je normálne rozdelenie,
3. je splnená podmienka homoskedasticity, čo znamená platnosť hypotézy:

$$H_0\colon\ \sigma_1^2=\sigma_2^2=\ldots=\sigma_k^2 \tag{3}$$

Podmienku nezávislosti náhodných výberov obyčajne posudzujeme logicky a zabezpečujeme vhodným výberom vzorky. Preto v praxi overujeme hlavne druhú a tretiu podmienku. Čo sa v skutočnosti stane ak skúmané skupiny nespĺňajú niektorý z týchto predpokladov? Sú závery metódy ANOVA platné? [Nan00] Aká robustná je táto metóda vo vzťahu k porušeniu základných predpokladov? Na tieto otázky nemôžeme získať odpoveď pomocou klasickej štatistickej teórie [Gla72], ale získame ich práve pomocou simulačnej Monte Carlo štúdie, v rámci ktorej budeme sledovať aktuálnu chybu I. druhu metódy ANOVA (p-hodnotu).

Konkrétne, v tomto príspevku budeme testovať hypotézu o zhode priemerov troch skupín

$$H_0\colon\ \mu_1=\mu_2=\mu_3=50 \tag{4}$$

oproti alternatívnej hypotéze, ktorá hovorí, že aspoň dva priemery nie sú zhodné. Pri návrhu simulačnej štúdie budeme postupovať podľa vyššie uvedených krokov 1-8.

- *Návrh otázok:*

Sú závery metódy ANOVA platné ak nie je splnená podmienka 2 alebo 3, alebo obe súčasne?

Aká robustná je táto metóda vo vzťahu k porušeniu základných predpokladov?

- *Návrh simulačnej štúdie:*

V reálnej Monte Carlo simulačnej štúdii musíme uvažovať všetky možnosti, ktoré môžu nastať. Ide o splnenie/nesplnenie podmienky normality údajov a zhodu/nezhodu populačného rozptylu jednotlivých skupín. V štúdii preto budeme zámerne manipulovať s rozptylmi rôznych populačných skupín, navrhneme vzorky pre populácie skupín pochádzajúce/nepochádzajúce z normálneho rozdelenia a otestujeme či jednotlivé skupiny majú zhodnú strednú hodnotu. Pre každú realizáciu zistíme či zhodu stredných hodnôt zamietame alebo nie (na základe *p*-hodnoty).

Uvažujme teda nasledujúce prípady:

- údaje pochádzajú z normálneho rozdelenia s populačným priemerom $\mu = 50$,
- údaje pochádzajú z rozdelenia, ktoré má šikmosť (skewness) rovnú 1,75 a strmosť (kurtosis) rovnú 3,75.
- V prípade zhodných rozptylov predpokladajme, že $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 10$.
- V prípade rôznych rozptylov predpokladajme, že $\sigma_1^2 = 10,\ \sigma_2^2 = 20,\ \sigma_3^2 = 40$.
- Pre jednoduchosť predpokladajme, že jednotlivé skupiny vzoriek majú rovnaký počet pozorovaní napr. 30.

Možnosti, ktoré môžu nastať sme zapísali do tabuľky 1:

	Normalita údajov		
Rozptyly		Splnená normalita	Nesplnená normalita
	Zhodné (10, 10, 10)	5000 opakovaní	5000 opakovaní
	Nezhodné (10, 20, 40)	5000 opakovaní	5000 opakovaní

Tabuľka 1: Schéma pre ANOVA Monte Carlo štúdiu

Pre každú bunku tabuľky spravíme 5000 realizácií, čiže celkovo získame 2*2*5000 =20000 realizácií simulačnej štúdie na výpočet chyby I. druhu metódy ANOVA.

- *Generovanie údajov:*

Dôležitú úlohu v Monte Carlo simulačných štúdiách má generovanie údajov. V našom príklade budeme generovať dva typy údajov.

V prípade generovania údajov pochádzajúcich z normálneho rozdelenia použijeme SAS náhodný generátor RANNOR pre hodnoty náhodnej premennej *X* pochádzajúce z normálneho rozdelenia N (0,1). Nakoľko potrebujeme generovať hodnoty z rozdelenia s daným μ a σ použijeme známu lineárnu transformáciu:

$$X_{new} = X * \sigma + \mu \qquad (5)$$

kde X_{new} je transformovaná hodnota premennej *X*, μ je želaná hodnota priemeru a σ je želaný rozptyl. Nová premenná X_{new} má rozdelenie $N(\mu,\sigma^2)$ (napríklad $N(50,10^2)$).

V prípade generovania údajov so želanou šikmosťou a strmosťou je vhodné použiť Fleishmanovu mocninnú transformačnú metódu (pozri v [Fle78] alebo v [Fan02]. Táto metóda používa polynomickú mocninnú transformáciu premennej *Z* s normálnym rozdelením do premennej so želanou šikmosťou a strmosťou. Fleishmanova polynomická mocninná transformácia má tvar:

$$Y = a + bZ + cZ^2 + dZ^3 \qquad (6)$$

kde *Y* je transformovaná premenná s požadovanou populačnou šikmosťou a strmosťou, *Z* je premenná s normálnym rozdelením *N*(0,1) a *a, b, c, d* sú koeficienty potrebné pre transformáciu z normálneho rozdelenia do nenormálneho rozdelenia s danou šikmosťou a strmosťou. Naviac platí, že *a=-c*.

Koeficienty *a, b, c, d* potrebné pre vyčíslenie transformácie boli tabelované pre niektoré vybrané dvojice šikmosti a strmosti [Fle78]. Pre účely tohto príspevku sme použili SAS/IML program na výpočet Fleishmanových koeficientov [Fan02], ktorý využíva Newton-Raphsonovu

iteračnú metódu. Pre šikmosť (skewness) rovnú 1,75 a strmosť (kurtosis) rovnú 3,75 vidíme hodnoty transformačných koeficientov v tabuľke 2:

The SAS System COEFFICEINTS OF B, C, D FOR FLEISHMAN'S POWER TRANSFORMATION Y = A + BX + CX^2 + DX^3, A = -C				
Result				
SKEWNESS	KURTOSIS	B	C	D
1,75	3,75	0,92966053	0,399497	-0,036466993

Tabuľka 2: Fleishmannove transformačné koeficienty

- *Implementácia kvantitatívnej (štatistickej) metódy*

Použijeme procedúru ANOVA, ktorú ponúka SAS/STAT softvér a nastavíme hodnotu premennej SIG na YES (ak hypotézu (4) zamietame) a na NO (ak hypotézu (4) nezamietame). Premenná NORMAL má hodnotu YES ak údaje pochádzajú z normálneho rozdelenia a hodnotu NO inak. Podobne premenná EQ_VAR má hodnotu YES ak skupiny vo vzorke majú zhodný rozptyl a NO ak nemajú. Ukážku SAS výstupu vidíme v tabuľke 3:

Row number	F	PROB	DF_MOD	SS_MOD	SIG	DF_ERR	SS_ERR	NORMAL	EQ_VAR
4	0,177	0,8385	2	3,26	NO	87	803,31	YES	YES
5	3,679	0,0293	2	91,521	YES	87	1082,02	YES	YES
7167	0,814	0,4465	2	42,997	NO	87	2297,93	YES	NO
7168	3,34	0,04	2	169,782	YES	87	2211,03	YES	NO
11414	0,645	0,5272	2	10,917	NO	87	736,29	NO	YES
11415	3,821	0,0257	2	71,626	YES	87	815,42	NO	YES
19999	3,22	0,0448	2	120,268	YES	87	1624,82	NO	NO
20000	0,805	0,4506	2	32,46	NO	87	1754,89	NO	NO

Tabuľka 3: Ukážka výstupu procedúry ANOVA v simulačnej štúdii Monte Carlo

- *Analyzujeme vyčíslené a zhromaždené štatistiky.*

Na záver utriedime výstup z procedúry ANOVA (SAS procedúrou SORT) a zistíme empirické rozdelenie počtu zamietnutí/nezamietnutí základnej hypotézy (4) ANOVY pre jednotlivé bunky tabuľky 1 (pomocou SAS procedúry FREQ). Výsledok vidíme v tabuľke 4:

The SAS System, The FREQ Procedure

NORMAL=NO, EQ_VAR=NO					NORMAL=YES, EQ_VAR=NO				
SIG	Frequency	Percent	Cumulative Frequency	Cumulative Percent	SIG	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NO	4679	93,58	4679	93,58	NO	4735	94,70	4735	94,70
YES	321	6,42	5000	100,00	YES	265	5,30	5000	100,00
NORMAL=NO, EQ_VAR=YES					NORMAL=YES, EQ_VAR=YES				
SIG	Frequency	Percent	Cumulative Frequency	Cumulative Percent	SIG	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NO	4765	95,30	4765	95,30	NO	4756	95,12	4756	95,12
YES	235	4,70	5000	100,00	YES	244	4,88	5000	100,00

Tabuľka 4: Výstup Monte Carlo simulačnej štúdie

- *Vyhodnotíme empiricky získané výsledky v Monte Carlo simulačnej štúdii.*

Pre každú bunku tabuľky 1 sme zistili percentuálny podiel zamietnutí hypotézy (4) na hladine významnosti 5% a výsledky sme zapísali do tabuľky 4. V simulačnej štúdii naše skupiny údajov mali rovnaký priemer a zámerne sme ovplyvňovali splnenie/nesplnenie predpokladov o normalite a zhode rozptylov metódy ANOVA, čiže hypotéza (4) by nemala byť zamietnutá. V prípade splnenia požiadavky zhody rozptylov však v 4,88% prípadov došlo k zamietnutiu hypotézy (4) pre normálne rozdelené údaje a v 4,70% prípadoch pre nenormálne rozdelené údaje. Chyby I. druhu sme sa dopustili v menej ako 5% prípadoch. Keď požiadavka o zhode rozptylov nebola dodržaná, hypotézu (4) sme zamietli vo viac ako v 5% prípadoch a teda dopustili sme sa chyby I. druhu častejšie. Môžeme preto usúdiť, že pre túto konkrétnu simulačnú štúdiu metóda ANOVA je citlivejšia na požiadavku zhody rozptylov ako na požiadavku normality údajov [Nan00].

Zoznam použitej literatúry

Fan02 Fan, X.- Felsovályi, Á.- Sivo, S.A.- Keenan, S. C.: *SAS for Monte Carlo Studies. A guide for Quantitative Researchers* . 1. vyd. North Carolina: SAS Institute Inc., 2002. ISBN 1-59047-141-5

Fle78 Fleishman, A.I.: *A method for Simulating Non-Normal Distributions.* In: Psychometrika, č. 43, 1978, s. 521-531.

Gla72 Glaas, G. V.- Peckham, P.D. - Sanders, J.R.: *Consequences of Failure to Meet Assumptions Underlying the Fixed-Effects Analysis of Variance and Covariance*. In: Review of Educational Research, č. 42, 1972, s. 237-288.

Nan00 Nánásiová, O.-Kalina, M. : *Problémy analýzy rozptylu*. In: Slovenská štatistika, súčasnosť a perspektívy:Zborník 10. Slovenskej štatistickej konferencie konanej v dňoch 10.-12.5.2000 v Smoleniciach.Bratislava: SŠDS,2000, s.121-125. ISBN 80-88946-06-9

Pac03 Pacáková, V. a kol.: *Štatistika pre ekonómov* . 1. vyd. Bratislava: IURA EDITION, 2003. ISBN 80-89047-74-2

Sta00 Stankovičová I.: *Ako robiť štatistiku v systéme SAS* . In: Výpočtová štatistika:Zborník 9. medzinárodného seminára konaného v dňoch 7.-8.12.2000 v Bratislave. Bratislava: SŠDS, 2000, s.74-78. ISBN 80-88946-09-3

Autor: RNDr. Mária Bohdalová
KIS FM UK
Odbojárov 10
820 05 Bratislava
mailto:maria.bohdalova@fm.uniba.sk

POROVNANIE AKAHIROVEJ APROXIMÁCIE S EXAKTNÝM VÝPOČTOM JEDNOSTRANNÝCH TOLERANČNÝCH ČINITEĽOV NORMÁLNEHO ROZDELENIA

Ivan Garaj

Abstrakt. V príspevku je uvedené porovnanie Akahirovej aproximácie s exaktným výpočetom jednostranných tolerančných činiteľov normálneho rozdelenia s neznámymi parametrami. Numerický výstup je uvedený v tabuľkách č. 1 až 3.

Kľúčové slová: Jednostranný tolerančný činiteľ; necentrálne t-rozdelenie; Cornishov-Fisherov rozvoj; Wallisova, Jennettova a Welchova, van Eedenova a Akahirova aproximácia jednostranných tolerančných činiteľov.

Úvod

Exaktný výpočet jednostranných tolerančných činiteľov k normálneho rozdelenia úzko súvisí s exaktným výpočtom kvantilov $t'_{\alpha}(v, \delta)$ necentrálneho t-rozdelenia, kde $v = n - 1$ sú stupne voľnosti a δ $(-\infty < \delta < \infty)$ je parameter necentrality tohto rozdelenia. Najrozsiahlejšie tabuľky kvantilov necentrálneho t-rozdelenia možno nájsť v monografii [1], kde sú vypočítané s presnosťou na 5 desatinných miest pre $\alpha =$ 0,01; 0,025; 0,05; 0,10; 0,20; 0,30; 0,70; 0,80; 0,90; 0,95; 0,975; 0,99, $v = 1\,(1)\,60$ a $\delta = 0{,}1\,(0{,}1)\,8{,}0$.

Exaktný výpočet jednostranných tolerančných činiteľov

V monografii [2] sú uvedené rozsiahle tabuľky jednostranných tolerančných činiteľov k vypočítané pre parameter necentrality $\delta = u_p\sqrt{n}$ podľa vzťahu

$$k = -\frac{t'_{\alpha}\left(n-1, -u_p\sqrt{n}\right)}{\sqrt{n}} = \frac{t'_{1-\alpha}\left(n-1, u_p\sqrt{n}\right)}{\sqrt{n}} \tag{1}$$

s presnosťou na 4 desatinné miesta (zaokrúhľovanie nahor) pre koeficienty spoľahlivosti $1-\alpha =$ 0,001; 0,005; 0,01; 0,025; 0,05; 0,10; 0,25; 0,50; 0,75; 0,90; 0,95; 0,975; 0,99; 0,995; 0,999; 0,9999, hodnoty $p =$ 0,525; 0,55(0,05) 0,70; 0,725; 0,75(0,05) 0,90; 0,91(0,01) 0,97; 0,975; 0,98; 0,99; 0,991(0,001) 0,999; 0,9999 a rozsahy výberov $n =$ 2(1) 200; 205(5) 300; 310(10) 400; 425(25) 1000; 1500; 2000(1000) 5000; 10000. V poslednom riadku (∞) sú hodnoty kvantilov u_p normovaného normálneho rozdelenia. Exaktný výpočet tolerančných činiteľov k podľa vzťahu (1) je pre veľké n enormne zdĺhavý.

Približný výpočet jednostranných tolerančných činiteľov

Existuje veľa aproximácií jednostranných tolerančných činiteľov [3], [4]. V práci [5] možno nájsť ich rozsiahle numerické porovnanie, no väčšina z nich sa pre veľkú chybu nedá použiť. Najznámejšia a doteraz najpoužívanejšia je Wallisova [6]. Pre veľké n je pomerne presná. Bola

odvodená na základe aproximácie štatistiky $\overline{X}+kS$ normálnym rozdelením $N\left(\mu+k\sigma,\ \frac{\sigma^2}{n}+\frac{\sigma^2 k^2}{2n-2}\right)$. Jednostranný tolerančný činiteľ k je daný vzťahom

$$k \approx \frac{u_p+\sqrt{u_p^2-AB}}{A} \tag{2}$$

kde $A=1-\frac{u_{1-\alpha}^2}{2(n-1)}$, $B=u_p^2-\frac{u_{1-\alpha}^2}{n}$ a u_p a $u_{1-\alpha}$ sú kvantily normovaného normálneho rozdelenia $N(0,1)$. Trochu lepšie výsledky dáva Jennettova a Welchova [7] aproximácia pomocou kvantilov necentrálneho t-rozdelenia

$$t'(v,\delta) \approx \frac{\delta b_v+u_\alpha\sqrt{b_v^2+\left(1-b_v^2\right)\left(\delta^2-u_\alpha^2\right)}}{b_v^2-u_\alpha^2\left(1-b_v^2\right)} \tag{3}$$

kde b_v je dané vzťahom

$$b_v=\frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}\sqrt{\frac{2}{v}} \qquad (v=n-1) \tag{4}$$

Pre malé δ je výhodná aproximácia van Eedena [4], ktorá bola odvodená pomocou Cornishovho-Fisherovho [8] rozvoja $(v=n-1)$

$$t'(v,\delta) \approx u_\alpha+\frac{u_\alpha^3+u_\alpha}{4v}+\frac{5u_\alpha^5+16u_\alpha^3+3u_\alpha}{96v^2}+$$
$$+\delta\left(1+\frac{2u_\alpha^2+1}{4v}+\frac{u_\alpha}{4v}\delta+\frac{4u_\alpha^4+12u_\alpha^2+1}{32v^2}+\frac{u_\alpha^3+4u_\alpha}{16v^2}\delta-\frac{u_\alpha^2-1}{24v^2}\delta^2-\frac{u_\alpha}{32v^2}\delta^3\right) \tag{5}$$

Ani o jednej z horeuvedených aproximácií sa nedá povedať, že dávajú vo všeobecnosti dobré výsledky. Tento nedostatok odstraňuje Akahirova [9] aproximácia, no porovnanie s exaktným výpočtom už v práci [5] nie je urobené. Kvantily necentrálneho t-rozdelenia $x=t'(v,\delta)$ sa nájdu riešením rovnice

$$\frac{b_v x-\delta}{\sqrt{1+x^2\left(1-b_v^2\right)}}=u_\alpha-\frac{x^3\left(u_\alpha^2-1\right)}{24\sqrt{\left[1+x^2\left(1-b_v^2\right)\right]^3}}\left(\frac{1}{v^2}+\frac{1}{4v^3}\right) \qquad (v=n-1) \tag{6}$$

kde b_v je dané vzťahom (4) a u_α je kvantil normovaného normálneho rozdelenia $N(0,1)$. Označme

$$g=\left(\frac{1}{v^2}+\frac{1}{4v^3}\right);\ \ c=b_v;\ \ h=1-b_v^2;\ \ r=g\left(u_\alpha-\frac{1}{u_\alpha}\right);\ \ \ s=\frac{c}{u_\alpha};\ \ \ m=\frac{\delta}{u_\alpha} \tag{7}$$

Rovnicu (6) možno potom prepísať do tvaru:

$$\left(1+h\,x^2\right)^3=\left[(s\,x-m)\left(1+h\,x^2\right)+r\,x^3\right]^2 \tag{8}$$

Vzťah (8) je vzhľadom k označeniu (7) algebraická rovnica šiesteho stupňa typu

$$a_0\, x^6 + a_1\, x^5 + a_2\, x^4 + a_3\, x^3 + a_4\, x^2 + a_5\, x + a_6 = 0 \tag{9}$$

s reálnymi koeficientami

$$\begin{aligned} a_0 &= -h^3 + r^2 + 2\, h\, r\, s + h^2\, s^2 \\ a_1 &= -2\, h\, r\, m - 2\, h^2\, s\, m \\ a_2 &= -3\, h^2 + 2\, r\, s + 2\, h\, s^2 + h^2\, m^2 \\ a_3 &= -2\, r\, m - 4\, h\, s\, m \\ a_4 &= -3\, h + s^2 + 2\, h\, m^2 \\ a_5 &= -2\, s\, m \\ a_6 &= m^2 - 1 \end{aligned} \tag{10}$$

Algebraická rovnica (9) bola numericky riešená v systéme Mathematica [10] využitím príkazu

$$FindRoot[a_0\, x^6 + a_1\, x^5 + a_2\, x^4 + a_3\, x^3 + a_4\, x^2 + a_5\, x + a_6 == 0,\quad \{x, x_0\}]$$

Za štartovací bod x_0 bol zvolený numerický výstup z Wallisovej aproximácie, vypočítanej podľa vzťahu (2).

Aproximácia (6) dáva dobré výsledky aj pre malé v. V tabuľkách č. 1 až 3 sú jednostranné tolerančné činitele k vypočítané exaktne na 8 desatinných miest pre parameter necentrality $\delta = u_p \sqrt{n}$ podľa vzťahu (1) a aj pomocou Akahirovej aproximácie (6). Z tabuliek vidno, že Akahirova aproximácia dáva solídne numerické výstupy. S rastúcim p rastie aj jej nepresnosť.

Tolerančné činitele majú veľa aplikácií pri štatistickom riadení kvality. Metódami štatistického riadenia kvality sa zaoberajú aj práce [11], [12], [13], [14], [15], [16], [17] a [18].

Tento príspevok bol vypracovaný s podporou grantovej agentúry VEGA v rámci projektu číslo 1/1247/04 Progresívne štatistické techniky a rozhodovanie v procese zlepšovania kvality.

Literatúra

[1] BAGUI, S. C. *CRC Handbook of Percentiles of Non-Central t-Distributions*. CRC Press, Boca Raton, Florida, 1993, 400 p. ISBN 0-8493-8669-1.

[2] GARAJ, I., JANIGA I. *Jednostranné tolerančné medze pre neznámu strednú hodnotu a rozptyl normálneho rozdelenia. One Sided Tolerance Limits of Normal Distribution for Unknown Mean and Variability*. Bratislava: Vydavateľstvo STU, 2005, 214 s. ISBN 80-227-2218-9.

[3] JÍLEK, M. *Statistické tolerančí meze*. Praha, SNTL, 1988, 275 s. (in Czech).

[4] EEDEN, van C. Some approximations to the percentage points of the non-central t-distribution. In *Revue de l´ Institut International de Statistique* 29, 1961, p. 4-31.

[5] SAHAI, H., OJEDA, M. M. A comparison of approximations to percentiles of the non-cental t-distribution. In *Revista Investigacion Operacional*, 2000, Vol. 21, No. 2, p.121-144.
[6] WALLIS, W. A. *Use of Variables in Acceptance Inspection for Percent Defective. Selected Techniques of Statistical Analysis* (EISENHART, C., HASTAY, M. W., WALLIS, W. A. eds.). New York, 1947, McGraw Hill Bokk Company, p. 7-93.
[7] JENNETT, W. T., WELCH, B. L. The control of proportion defective as judged by a single quality characteristic varying on a continuous scale. In *Journal of the Royal Statistical Society*, B 6, 1939, p. 80-88.
[8] CORNISH, E. A., FISHER, R. A. Moments and cumulants in the specification of distributions. In *Revue de l´ Institut International de Statistique*, 5, 1937, p. 307-320.
[9] AKAHIRA, M. A higher order aproximation to a percentage point of the non-central t-distribution. In *Communications in Statistics*, Part B: Simulcition and Computation, 1995, 24(3), p. 595-605.
[10] WOLFRAM, S. *The Mathematica Book*. 3rd ed. Wolfram Media/Cambridge University Press, 1996. 1403 p. ISBN 0-521-58889-8.
[11] GARAJ, I., JANIGA I. *Dvojstranné tolerančné medze pre neznámu strednú hodnotu a rozptyl normálneho rozdelenia*. Bratislava, Vydavateľstvo STU, 2002, 147 s. ISBN 80-227-1779-7.
[12] GARAJ, I., JANIGA I. *Dvojstranné tolerančné medze normálnych rozdelení s neznámymi strednými hodnotami a s neznámym spoločným rozptylom. Two Sided Tolerance Limits of Normal Distributions with Unknown Means and Unknown Common Variability.* Bratislava, Vydavateľstvo STU, 2004, 218 s. ISBN 80-227-2019-4.
[13] ODEH, R.E., OWEN, D.B. *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*. New York, Marcel Dekker, 1980, 316 p. ISBN 0-8247-6944-9.
[14] GARAJ, I. Preberací plán meraním pri daných jednostranných tolerančných medziach. In *Statistika*, 10, 1992, s. 431-437 (in Slovak).
[15] JANIGA, I., GARAJ, I. On One-Sided Tolerance Intervals of Normal Distribution with Unknown Parameters. In *Journal of the Applied Mathematics, Statistics and Informatics* (JAMSI). ISBN 80-89034-95-0, 2005, vol. 1, no. 1, p. 87-100.
[16] JANIGA, I., GARAJ, I. On Exact Computation of One-sided Tolerance Factors of Normal Distributions. In FORUM STATISTICUM SLOVACUM. ISSN 1336-7420, 1/2005, p. 74-79.
[17] JANIGA, I., MIKLÓŠ, R. Statistical tolerance intervals for a normal distribution. In *Measurement Science Review*. ISSN 13, 2001, vol. 1, no. 1, p. 29-32.
[18] TEREK, M., HRNČIAROVÁ, Ľ. *Štatistické riadenie kvality*. Vydavateľstvo IURA EDITION, 2004. 234 s. ISBN 80-89047-97-1. (in Slovak).

Kontaktná adresa autora

RNDr. Ivan Garaj, CSc.,
Katedra matematiky, FCHPT STU,
Radlinského 9, 812 37 Bratislava
Tel. 00421-7-59325297,
E-mail: ivan.garaj@stuba.sk

Tabuľka č. 1: Jednostranné tolerančné činitele *k*

1-α = 0,025				
p	*n*	P(resne)	A(kahira)	| P-A |
0,75	100	0,46899029	0,46899023	0,00000006
	150	0,50525981	0,50525976	0,00000005
	200	0,52713577	0,52713573	0,00000004
	250	0,54218523	0,54218519	0,00000004
	300	0,55336094	0,55336091	0,00000003
	400	0,56914859	0,56914856	0,00000003
	500	0,57999304	0,57999302	0,00000002
0,90	100	1,03971018	1,03970684	0,00000334
	150	1,08110744	1,08110524	0,00000220
	200	1,10635337	1,10635178	0,00000159
	250	1,12383984	1,12383861	0,00000123
	300	1,13688705	1,13688606	0,00000099
	400	1,15540723	1,15540654	0,00000069
	500	1,16818825	1,16818771	0,00000054
0,95	100	1,37304447	1,37303448	0,00000999
	150	1,41910711	1,41910085	0,00000626
	200	1,44729787	1,44729338	0,00000449
	250	1,46686693	1,46686359	0,00000334
	300	1,48149058	1,48148793	0,00000265
	400	1,50228098	1,50227916	0,00000182
	500	1,51665062	1,51664926	0,00000136
0,975	100	1,65906317	1,65904434	0,00001883
	150	1,70976512	1,70975363	0,00001149
	200	1,74085677	1,74084880	0,00000797
	250	1,76246656	1,76246056	0,00000600
	300	1,77862925	1,77862452	0,00000473
	400	1,80162818	1,80162496	0,00000322
	500	1,81753816	1,81753578	0,00000238
0,99	100	1,98912392	1,98909210	0,00003182
	150	2,04569204	2,04567297	0,00001907
	200	2,08043434	2,08042121	0,00001313
	250	2,10460479	2,10459502	0,00000977
	300	2,12269507	2,12268741	0,00000766
	400	2,14845490	2,14844971	0,00000519
	500	2,16628700	2,16628318	0,00000382
0,995	100	2,21275061	2,21270881	0,00004180
	150	2,27352303	2,27349819	0,00002484
	200	2,31087506	2,31085805	0,00001701
	250	2,33687329	2,33686067	0,00001262
	300	2,35633802	2,35632816	0,00000986
	400	2,38406444	2,38405778	0,00000666
	500	2,40326433	2,40325944	0,00000489
0,999	100	2,67182419	2,67176017	0,00006402
	150	2,74163990	2,74160231	0,00003759
	200	2,78459495	2,78456938	0,00002557
	250	2,81451296	2,81449407	0,00001889
	300	2,83692291	2,83690828	0,00001463
	400	2,86886034	2,86885046	0,00000988
	500	2,89098679	2,89097955	0,00000724
0,9999	100	3,23046676	3,23037426	0,00009250
	150	3,31179793	3,31174410	0,00005383
	200	3,36187636	3,36183993	0,00003643
	250	3,39677290	3,39674608	0,00002682
	300	3,42292125	3,42290041	0,00002084
	400	3,46019966	3,46018571	0,00001395
	500	3,48603566	3,48602547	0,00001019

Tabuľka č. 2: Jednostranné tolerančné činitele *k*

1-α = 0,90				
p	*n*	P(resne)	A(kahira)	| P-A |
0,75	100	0,82496417	0,82496372	0,00000045
	150	0,79589845	0,79589823	0,00000022
	200	0,77891936	0,77891923	0,00000013
	250	0,76747238	0,76747229	0,00000009
	300	0,75909265	0,75909258	0,00000007
	400	0,74742721	0,74742717	0,00000004
	500	0,73952943	0,73952940	0,00000003
0,90	100	1,47006157	1,47005575	0,00000582
	150	1,43284190	1,43283897	0,00000293
	200	1,41127707	1,41127526	0,00000181
	250	1,39681264	1,39681139	0,00000125
	300	1,38626197	1,38626104	0,00000093
	400	1,37162791	1,37162732	0,00000059
	500	1,36175575	1,36175534	0,00000041
0,95	100	1,86125165	1,86123889	0,00001276
	150	1,81819740	1,81819092	0,00000648
	200	1,79332398	1,79331995	0,00000403
	250	1,77667041	1,77666760	0,00000281
	300	1,76453828	1,76453618	0,00000210
	400	1,74773233	1,74773100	0,00000133
	500	1,73640931	1,73640838	0,00000093
0,975	100	2,20251235	2,20249185	0,00002050
	150	2,15403332	2,15402284	0,00001048
	200	2,12607454	2,12606800	0,00000654
	250	2,10737553	2,10737095	0,00000458
	300	2,09376366	2,09376024	0,00000342
	400	2,07492249	2,07492032	0,00000217
	500	2,06223787	2,06223635	0,00000152
0,99	100	2,60090282	2,60087205	0,00003077
	150	2,54581958	2,54580379	0,00001579
	200	2,51409750	2,51408759	0,00000991
	250	2,49290047	2,49289353	0,00000694
	300	2,47747985	2,47747466	0,00000519
	400	2,45614862	2,45614533	0,00000329
	500	2,44179656	2,44179424	0,00000232
0,995	100	2,87290442	2,87286621	0,00003821
	150	2,81318983	2,81317017	0,00001966
	200	2,77882546	2,77881310	0,00001236
	250	2,75587316	2,75586451	0,00000865
	300	2,73918090	2,73917442	0,00000648
	400	2,71609802	2,71609391	0,00000411
	500	2,70057230	2,70056940	0,00000290
0,999	100	3,43506189	3,43500781	0,00005408
	150	3,36555386	3,36552595	0,00002791
	200	3,32559638	3,32557880	0,00001758
	250	3,29892619	3,29891387	0,00001232
	300	3,27953912	3,27952989	0,00000923
	400	3,25274247	3,25273659	0,00000588
	500	3,23472721	3,23472306	0,00000415
0,9999	100	4,12387389	4,12380015	0,00007374
	150	4,04208916	4,04205101	0,00003815
	200	3,99511323	3,99508917	0,00002406
	250	3,96377468	3,96375779	0,00001689
	300	3,94100233	3,94098966	0,00001267
	400	3,90953811	3,90953005	0,00000806
	500	3,88839248	3,88838679	0,00000569

Tabuľka č. 3: Jednostranné tolerančné činitele *k*

1-α = 0,999				
p	*n*	P(resne)	A(kahira)	|P-A|
0,75	100	1,06123298	1,06127084	0,00003786
	150	0,98201786	0,98203225	0,00001439
	200	0,93689444	0,93690194	0,00000750
	250	0,90691696	0,90692156	0,00000460
	300	0,88518730	0,88519042	0,00000312
	400	0,85522680	0,85522851	0,00000171
	500	0,83512612	0,83512722	0,00000110
	600	0,82046024	0,82046100	0,00000076
0,90	100	1,77498379	1,77524888	0,00026509
	150	1,67066948	1,67078452	0,00011504
	200	1,61192396	1,61198919	0,00006523
	250	1,57318374	1,57322626	0,00004252
	300	1,54525079	1,54528099	0,00003020
	400	1,50694835	1,50696615	0,00001780
	500	1,48139187	1,48140379	0,00001192
	600	1,46281842	1,46282705	0,00000863
0,95	100	2,21427388	2,21477206	0,00049818
	150	2,09266779	2,09289172	0,00022393
	200	2,02443846	2,02456784	0,00012938
	250	1,97955112	1,97963710	0,00008598
	300	1,94724196	1,94730374	0,00006178
	400	1,90301776	1,90305480	0,00003704
	500	1,87356279	1,87358788	0,00002509
	600	1,85218338	1,85220171	0,00001833
0,975	100	2,60000221	2,60073803	0,00073582
	150	2,46251183	2,46284899	0,00033716
	200	2,38553678	2,38573426	0,00019748
	250	2,33496654	2,33509869	0,00013215
	300	2,29860324	2,29869881	0,00009557
	400	2,24888134	2,24893916	0,00005782
	500	2,21579896	2,21583839	0,00003943
	600	2,19180439	2,19183334	0,00002895
0,99	100	3,05238215	3,05341639	0,00103424
	150	2,89568707	2,89616815	0,00048108
	200	2,80811221	2,80839716	0,00028495
	250	2,75064333	2,75083490	0,00019157
	300	2,70935232	2,70949153	0,00013921
	400	2,65293924	2,65302407	0,00008483
	500	2,61543598	2,61549411	0,00005813
	600	2,58825102	2,58829386	0,00004284
0,995	100	3,36219880	3,36344311	0,00124431
	150	3,19208869	3,19267172	0,00058303
	200	3,09709862	3,09744551	0,00034689
	250	3,03479836	3,03503231	0,00023395
	300	2,99005386	2,99022428	0,00017042
	400	2,92894765	2,92905185	0,00010420
	500	2,88834107	2,88841264	0,00007157
	600	2,85891522	2,85896805	0,00005283
0,999	100	4,00426774	4,00595094	0,00168320
	150	3,80587461	3,80667153	0,00079692
	200	3,69523115	3,69570837	0,00047722
	250	3,62272330	3,62304660	0,00032330
	300	3,57067783	3,57091414	0,00023631
	400	3,49964341	3,49978860	0,00014519
	500	3,45246743	3,45256749	0,00010006
	600	3,41829550	3,41836955	0,00007405
0,9999	100	4,79322475	4,79544286	0,00221811
	150	4,55946991	4,56052832	0,00105841
	200	4,42923109	4,42986797	0,00063688
	250	4,34393461	4,34436753	0,00043292
	300	4,28273663	4,28305385	0,00031722
	400	4,19924833	4,19944393	0,00019560
	500	4,14382647	4,14396160	0,00013513
	600	4,10369456	4,10379475	0,00010019

Porovnanie modelov CAPM a APT v slovenských podmienkach

Rudolf Gavliak

Abstract

This paper compares the explanation power of the wide used CAPM model with the less known APT model, which is using more explanatory variables to determine the asset return. The nature of Slovak capital market allows only a few case studies and it is not possible to derive a general relation equation. In this paper the two mentioned models are estimated to fit Slovnaft share monthly returns.

The CAPM model realized quite well in the Slovnaft share case, but when testing other shares returns the explanatory power decreases rapidly. This is probably caused by great weight of Slovnaft share in Slovak Stock Exchange index.

In this case the APT model explained even less dependent variable variability than the simple model. This is in conflict with theoretical expectations and empirical findings. The reason is that higher return is empirically not connected with increase of macroeconomic variables, but with their variance. Higher volatility means higher risk and higher risk premium demanded. Recent studies tested economic importance of income (industrial production) uncertainty to shares returns value. A generalized autoregressive conditional heteroskedasticity measure of income uncertainty is often used as a priced factor in a model of the arbitrage pricing theory. But in this case the presence of ARCH was proven only in long-term interest rates. The variance was described with GARCH (1,1) model.

The GARCH term performed well as a pricing factor in CAPM and also in APT model and helped to increase the explained dependent variable variance. This is probably not only this share case, because similar results received with SES Tlmače share, which is not traded so often and also the relative weight in SAX index is much lower.

Keywords: CAPM, APT, Uncertainty, GARCH, Shares Volatility models, Time series, Heteroskedasticity, Appraisal.

Úvod

Model CAPM meria citlivosť výnosu sledovaného aktíva na rozdiely výnosu trhového a bezrizikového aktíva. Tento jednoduchý model tvaru (1) našiel uplatnenie v mnohých finančných aplikáciách.

$$E(r_i) = r_f + \beta \cdot (r_m - r_f) \tag{1}$$

Beta faktor CAPM modelu je relatívnou mierou rizika. Táto citlivosť na riziko ale môže pozostávať z rôznych faktorov. Napríklad v skupine aktív s rovnakou citlivosťou na zmenu trhového výnosu môžu byť aktíva s nízkou citlivosťou na zmenu inflácie a naopak vysokou produkčnou citlivosťou a naopak. Beta faktor neumožňuje merať druh rizika na ktorý je dané aktívum citlivé.

Ďalším problémom modelu CAPM je, že na výpočet trhového výnosu sa používajú burzové indexy, ktoré i keď často obsahujú značné množstvo aktív, nie sú optimalizované, čo potvrdzujú viaceré štúdie. Dôsledkom je, že aktíva s rovnakým beta faktorom dosahujú rôzne očakávané výnosy. Za týchto podmienok nie je beta faktor spoľahlivým deskriptorom očakávaného výnosu.

Výnos aktíva je možné zapísať ako súčet troch častí výnosu. Skutočný výnos pozostáva z očakávaného výnosu, resp. strednej hodnoty výnosu. Ďalším faktorom, ktorý vplýva na výnos aktíva je citlivosť na zmenu výšky systematického rizika. Tretím faktorom je nesystematické (špecifické riziko), ktoré je možné odstrániť pomocou diverzifikácie v prípade „dobre diverzifikovaného portfólia“. Táto skutočnosť za dá zapísať nasledujúcim spôsobom:

$$r_i = E(r_i) + \sum_{j=1}^{m} b_{ij} \cdot f_j + e_i \tag{2}$$

kde

$E(r_i)$ je očakávaná hodnota aktíva za predpokladu diverzifikácie špecifického rizika,

b_{ij} je citlivosť *i* – tého aktíva na *j* – ty fundamentálny faktor,

f_j je prémia za zvýšenie rizika spôsobené fundamentálnym faktorom,

e_i je riziková prirážka za špecifické riziko *i* – tého faktora.

Je možné povedať, že prirážka za špecifické riziko predstavuje hodnoty reziduí viacnásobného regresného modelu. V prípade odstránenia špecifického rizika diverzifikáciou spĺňajú reziduá špecifikáciu bieleho šumu, v prípade nenulového špecifického rizika platí: $E(\mathbf{e}) \neq \mathbf{0}$.

Chen (1986) považuje za vysvetľujúce faktory zmeny výnosov trhových aktív neanticipovanú infláciu, prírastok priemyselnej produkcie, rozdiel medzi dlhodobými a krátkodobými úrokovými mierami a rizikovú prirážku (meranú rozdielom medzi výnosom dlhopisov s vysokým a nízkym ratingom).
Neanticipovaná inflácia bola vypočítaná ako rozdiel medzi hodnotou reziduí procesu AR(1) nameranej inflácie a skutočnou infláciou. Neanticipovaná zmena výstupu je stotožnené s reziduami autoregresného procesu prvého rádu (*AR(1)*) priemyselnej produkcie. Rozdiel medzi cenou štátnych a korporátnych dlhopisov som považoval za mieru rizikovej prirážky medzi štátnymi a podnikateľskými cennými papiermi. Mierou rozdielu medzi dlhodobými a krátkodobými úrokovými mierami bude rozdiel vo výnosoch štátnych dlhopisov a štátnych pokladničných poukážok.
Nestabilná miera rastu spôsobuje ťažkosti pri plánovaní investícií a spotreby. Podmienený rozptyl priemyselnej produkcie by sa mal vyvíjať v súlade so zvýšeným rizikom.
Volatilita vysvetľujúcich premenných, ktorá predstavuje riziko zmeny vo výške ekonomického prostredia bude modelovaná modelom zovšeobecnenej podmienenej autoregresnej heteroskedasticity (GARCH). Modely GARCH sú vhodné vzhľadom na ich schopnosť dobre popísať podmienený rozptyl pri malom počte parametrov.

Modely autoregresnej podmienenej heteroskedasticity (ARCH) boli navrhnuté za účelom predpovede podmieneného rozptylu. Rozptyl závislej premennej (kurz) je modelovaný ako funkcia minulých hodnôt. ARCH modely boli zovšeobecnené do podoby všeobecných autoregresných heteroskedasticitných modelov (GARCH). Modely GARCH majú dva parametre, ktoré sa uvádzajú v zátvorke za označením modelu. Hodnota prvého parametra definuje hĺbku pamäte procesu druhých mocnín reziduí. Hodnota druhého parametra hovorí o hĺbke pamäte autoregresného procesu minulej variability. Kde pre podmienený rozptyl platí vzťah:

$$\sigma_t^2 = \omega + \alpha_1 \left(y_{t-1} - \beta x_{t-1}^{'} \right)^2 + \alpha_2 \sigma_{t-1}^2 \qquad (3)$$

Tento vzťah počíta odhad volatility ako súčet dlhodobého priemeru volatility a váženého priemeru predpoveď volatility z predchádzajúceho obdobia, a skutočnej volatility nameranej v predchádzajúcom období. V prípade, že cena aktíva (underlying) neočakávane poklesla, alebo vzrástla, potom vzrastie odhad budúcej volatility. Tento model vyhovuje, často pozorovanému efektu, kedy obdobie vysokých výnosov je nasledované obdobím s ešte vyšším výnosom a naopak (*volatility clustering effect*).
Je zrejmé, že model GARCH (1,1) má tri parametre $(\alpha_1, \alpha_2, \omega)$, ktoré je potrebné odhadnúť. Existujú taktiež modifikácie modelu GARCH (1,1), pri ktorých je parameter ω nahradený

výrazom $\omega = V(1-\alpha_1-\alpha_2)$, kde V je dlhodobý rozptyl. Táto modifikácia sa nazýva cielenie rozptylu (*variance targeting*) (Zmeškal, 2004, s. 177).
Vysvetľujúce makroekonomické premenné sme podrobili testu ARCH založenom na Lagrangeových multiplikátoroch. V prípade zamietnutia nulovej hypotézy predpokladáme zotrvačnosť vo výške rozptylu a v tomto prípade je možné a vhodné modelovať podmienený rozptyl pomocou GARCH modelov.

Dáta a metodika

Skúmali sme závislosť medzi výnosom ceny akcie Slovnaftu, ktorá je jednou z mála aktívne obchodovaných na slovenskom kapitálovom trhu, a percentuálnou zmenou indexu SAX. Vzhľadom na vysokú váhu analyzovanej akcie predpokladáme silný lineárny vzťah medzi týmito premennými. Odhadnutý tvar modelu je nasledovný:

$$\begin{aligned}(R_A - R_F) &= -0.3805 \; + \; 0.2610 * (R_M - R_F) \\ &\quad (-17.97) \qquad (6.90) \quad R^2_{adj.} = 0.45\end{aligned} \tag{4}$$

kde
R_A je výnos z akcie Slovnaftu ~ $\ln(P_t / P_{t-1})$,
R_M je percentuálna medzimesačná zmena indexu SAX,
R_F je výnos z 10 – ročných štátnych dlhopisov.
Hodnoty v zátvorkách sú hodnoty *t – štatistiky*, ktoré sú významné na všetkých bežných hladinách významnosti. Odhadnutý vzťah (4) prevedieme na tvar (2):

$$R_A = (R_F - 0.3805) + 0.2610 * (R_M - R_F) \tag{5}$$

Tento jednoduchý jednofaktorový model sa používa nielen na odhad budúcich výnosov, ale aj na odhad nákladov vlastného kapitálu, ktorý sa využíva pri analýze a oceňovaní podnikov. Predpokladom pre použitie sú reziduá spĺňajúce vlastnosti bieleho šumu. V podmienkach slovenského kapitálového trhu je ale dosiahnutie týchto predpokladov problematické. Z hodnoty indexu determinácie sa dá určiť, že nezávislá premenná nevysvetľuje ani polovicu variability nezávislej premennej.
V ďalšom kroku prijmeme predpoklad *Chena* (1986) o vzťahu medzi určenými makroekonomickými veličinami a výnosov aktív obchodovaných na kapitálovom trhu. Odhadli sme preto lineárny model, ktorý kvantifikuje vzťah medzi výnosom akcie Slovnaftu a neanticipovanou infláciou, neanticipovanou zmenou priemyselnej produkcie, rizikovou prirážkou súkromného sektora voči štátu a rozdielu medzi krátkodobými a dlhodobými úrokovými mierami. Model tohto vzťahu má nasledovný tvar:

$$\begin{aligned}R_A - R_F &= -0.4886 + 0.0432 INFL_{AR} - 0.0109 IP_{AR_{t-1}} + 0.1021 risk - 0.2525 dif \\ &\quad (-14.81) \qquad (2.07) \qquad (-1.84) \qquad (2.07) \qquad (-2.17) \quad R^2_{\text{adj.}} = 0.29\end{aligned} \tag{6}$$

kde
R_A je výnos z akcie Slovnaftu ~ $\ln(P_t / P_{t-1})$,
R_F je výnos z 10 – ročných štátnych dlhopisov.

$INFL_{AR}$ je rozdiel medzi infláciou v dvoch nasledujúcich mesiacoch,

$IP_{AR_{t-1}}$ je rozdiel v priemyselnej produkcii v dvoch nasledujúcich mesiacoch posunuté o jedno obdobie do minulosti,

risk je rozdiel vo výnosoch podnikových a štátnych dlhopisov,

dif je rozdiel v dlhodobých a krátkodobých úrokoch.

Podľa hodnôt *t – štatistík* (v zátvorke) sú na 5 % hladine významnosti nenulové všetky koeficienty s výnimkou parametra pri premennej vyjadrujúcej neanticipovanú zmenu priemyselnej produkcie. Táto premenná nie je teda vhodná na vysvetlenie variability závislej premennej. V rozpore s teoretickým predpokladom je aj nízka hodnota indexu determinácie.

V ďalšom kroku sme testovali jednotlivé vysvetľujúce premenné na prítomnosť autoregresnej podmienenej heteroskedasticity (ARCH). Prítomnosť efektu „zhlukovania rozptylu“ sa potvrdila iba v prípade dlhodobých úrokových mier na medzibankovom trhu. Prítomnosť ARCH sa nepotvrdila v prípade priemyselnej produkcie, čo je v rozpore s očakávaniami.

Zvýšená volatilita úrokových mier by ale takisto mohla vysvetľovať zmeny cien aktív obchodovaných na kapitálovom trhu, preto odhadnem GARCH model, popisujúci podmienený rozptyl dlhodobých úrokových mier. Odhadnutý model má nasledujúcu rovnicu strednej hodnoty:

$$\begin{matrix} R_t = \text{-}0.1526 + 1.0149\mathrm{R}_{\text{t-1}} + e_t \\ (\text{-}0.97) \quad (42.06) \end{matrix} \tag{7}$$

Koeficient pri oneskorenej dlhodobej úrokovej miere naznačuje explozívny charakter autoregresného procesu. Druhá mocnina reziduí rovnice strednej hodnoty (*ARCH term*) sa používa na predpoveď volatility. Odhad volatility je formulovaný nasledujúcim spôsobom:

$$\begin{matrix} \sigma_t^2 = 0.002313 - 0.0415 e_t^{\,2} + 1.056\sigma_{t-1}^2 \\ (1.96) \quad (\text{-}4.22) \quad (37.5) \end{matrix} \tag{8}$$

Dlhodobá zabudovaná volatilita je na úrovni 0,23 % mesačne. Vysoká hodnota koeficientu pri volatilite predchádzajúceho obdobia (*GARCH term*) hovorí o zotrvačnosti vo výške volatility a opačne nízky a záporný koeficient pri rozptyle reziduí pôsobí ako tlmiaci faktor.

Teoretický predpoklad hovorí o efekte zotrvačnosti vo volatilite (*volatility persistence*) v prípade, že súčet posledných parametrov sa rovná jednej. Waldov test nemôže zamietnuť nulovú hypotézu v tvare: $\alpha_1 + \alpha_2 = 1$. Z toho vyplýva, že pri zvýšení volatility dlhodobých úrokových mier, pokles volatility je iba pozvoľný.

Modelovaná volatilita úrokových mier sa navyše ukázala ako vhodná vysvetľujúca premenná výnosu oboch analyzovaných akcií Slovnaftu aj SES Tlmače.

Záver

Pre jednoduchosť uvádzam iba modifikovaný model arbitrážneho oceňovania pre akciu Slovnaftu, ktorý zahrňuje ako vysvetľujúcu premennú aj GARCH model volatility dlhodobých úrokových mier.

$$\begin{matrix} R_A - R_F = \text{-}0.7780 + 0.0593 INFL_{AR} - 0.0025 IP_{AR_{t-1}} - 0.0297 risk - 0.3544 dif + 3.8489\sigma^2 \\ (\text{-}13.99) \quad (2.39) \quad (\text{-}0.52) \quad (\text{-}0.76) \quad (\text{-}3.89) \quad (5.81) \quad \mathrm{R}^2_{\text{adj.}} = 0.57 \end{matrix} \tag{9}$$

kde

R_A je výnos z akcie Slovnaftu ~ $\ln(P_t / P_{t-1})$,

R_F je výnos z 10 – ročných štátnych dlhopisov.

$INFL_{AR}$ je rozdiel medzi infláciou v dvoch nasledujúcich mesiacoch,

$IP_{AR_{t-1}}$ je rozdiel v priemyselnej produkcie v dvoch nasledujúcich mesiacoch posunuté o jedno obdobie,

risk je rozdiel vo výnosoch podnikových a štátnych dlhopisov,

dif je rozdiel v dlhodobých a krátkodobých úrokoch,

σ^2 je odhad podmienenej volatility ročných úrokových sadzieb na medzibankovom trhu.

Vysoká hodnota koeficientu volatility dlhodobých úrokových mier naznačuje výrazný vplyv volatility dlhodobých úrokových mier na výnos akcie. Nárast podmieneného rozptylu úrokových mier spôsobil v minulosti ešte vyšší nárast výnosu sledovanej akcie. Opačné znamienko, oproti teoretickým predpokladom je možné zdôvodniť poklesom úrokových mier v sledovanom období (volatilita spôsobená poklesom úrokových mier), čo bolo spôsobené štandardizáciou ekonomických podmienok v Slovenskej republike a sprevádzané vyšším záujmom o investovanie na kapitálovom trhu.

Nízke hodnoty *t – štatistiky* koeficientov neanticipovanej zmeny priemyselnej produkcie a rizikovej prémie korporátnych dlhopisov zamieta hypotézu o vysvetľovacej schopnosti výnosu akcie Slovnaftu týchto premenných. Potvrdil som vzťah medzi volatilitou úrokových mier a výnosom konkrétnych akcií, čo naznačuje možnosť modelovať výnos akcií skôr pomocou modelov volatility ako strednej hodnoty.

Literatúra

BERNANKE, B.1983. Irreversibility, Uncertainty and Cyclical Investment. In: Quarterly Journal of Economics, Vol. 98, No 1, 1983. p. 85–106.

DHANKAR, R. J. 2005. Arbitrage pricing theory and the capital asset pricing model-Evidence from the Indian stock market. In: Journal of Financial Management and Analysis, Vol. 18, No 1, 2005. p. 14–27.

CHEN, N. F. - ROSS, S. A.- ROLL, R. 1986. Economic Forces and the Stock Market. In: Journal of Business, No 59, 1986. p. 383–403.

MCKIERNAN, B.1997. Uncertainty and the Arbitrage Pricing Theory. In: American Economic Journal, Vol. 25, No 3, 1997. p. 307–311.

MELICHERČÍK, I. - OLŠÁROVÁ, L.- ÚRADNÍČEK, V. 2005. Economic Kapitoly z finančnej matematiky 1. Zvolen: Bratia Sabovci, 2005. ISBN 80-89029-88-4.

ROSS, S. A. 1976. The Arbitrage pricing theory of Capital Assets Pricing. In: Journal of Economic Theory, 1976. p. 341–360.

ROSS, S. A.- ROLL, R. 1980. An Empirical Investigation of the Arbitrage Pricing Theory. In: The Journal of Finance, December, 1980. p. 1073–1104.

ZMEŠKAL, Z. A KOL. 2004. Finanční modely. Praha : EKOPRESS, 2004. ISBN 80-86119-87-4.

Kontaktná adresa

Ing. Rudolf Gavliak
Katedra kvantitatívnych metód a informatiky
Ekonomická fakulta UMB Banská Bystrica
Tajovského 10
974 24 Banská Bystrica
rudolf.gavliak@umb.sk

Využitie štatistického testovania na zhodnotenie vplyvu vybraných nástrojov marketingovej komunikácie na spotrebiteľa

Ivana Hroncová

Abstract: The aim of the article is to find the relation between the selected variables and then to estimate the consequence of the selected marketing communication implements in term of its influence to consumer. Together we find the selected implements consequence for the enterprises functioning in the dairy market.

Predložený príspevok prezentuje štatistické testovanie nezávislostí nominálnych premenných. Testovanie sme realizovali medzi výškou mesačných voľných disponibilných prostriedkov na osobu v domácnosti v Sk a výškou mesačných výdavkov na mliečne výrobky a medzi intenzitou vplyvu vybraných nástrojov marketingovej komunikácie, konkrétne z oblasti reklamy, a výškou mesačných výdavkov na mliečne výrobky.

Cieľom príspevku je zistiť, či existujú závislosti medzi uvedenými premennými a na základe toho zhodnotiť význam vybraných nástrojov marketingovej komunikácie z hľadiska ich vplyvu na spotrebiteľa a zároveň zistiť význam uvedených nástrojov marketingovej komunikácie pre podniky pôsobiace v odvetví mlieka a mliečnych výrobkov.

Nasledujúce tabuľky a grafy sú súčasťou tohto štatistického testovania zvolených premenných.

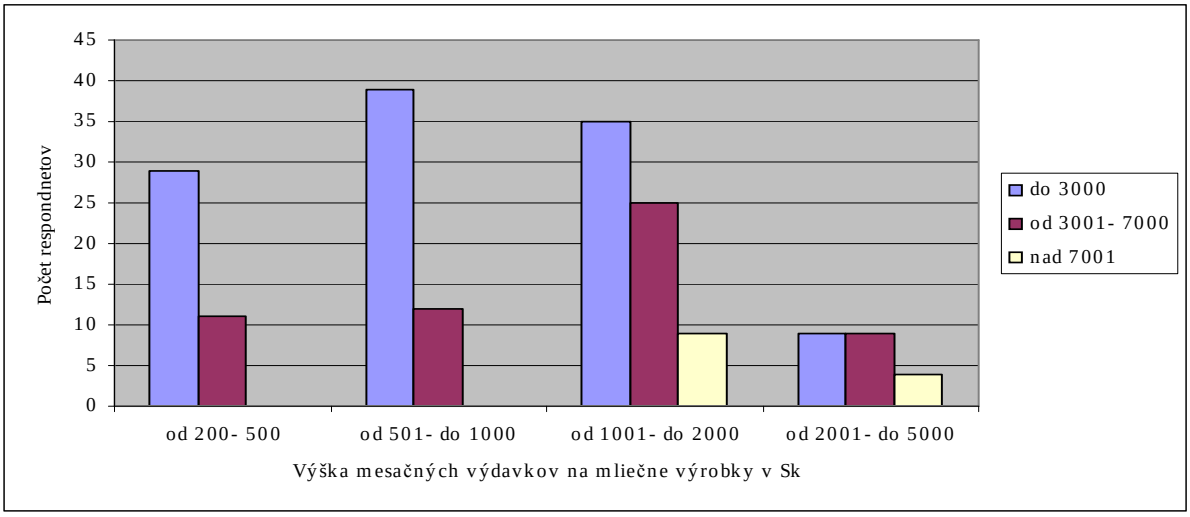
Tabuľka 1 Závislosť medzi výškou mesačných voľných disponibilných prostriedkov na osobu v domácnosti a výškou mesačných výdavkov na mliečne výrobky

Voľné disponibilné mesačné prostriedky na osobu v domácnosti v Sk	Výška mesačných výdavkov na mliečne výrobky v Sk				
	od 200- 500	od 501- do 1000	od 1001- do 2000	od 2001- do 5000	Celkový súčet respondentov
do 3000	29	39	35	9	112
od 3001- 7000	11	12	25	9	57
nad 7001	-	-	9	4	13
celkový súčet respondentov	40	51	69	22	182

Prameň: vlastný výskum

Pri testovaní závislosti medzi výškou mesačných voľných disponibilných prostriedkov na osobu v domácnosti v Sk a výškou mesačných výdavkov na mliečne výrobky je chí - kvadrát testovacia štatistika 21,43, kritická hodnota je 12,59 a miera tesnosti závislosti je 0,32. Na základe toho možno konštatovať, že medzi sledovaným premennými existuje silná závislosť. To znamená, že na základe výsledku uvedeného testovania vyplýva, že čím je výška mesačných voľných disponibilných prostriedkov na osobu v domácnosti vyššia, tým sú vyššie aj mesačné výdavky na mliečne výrobky.

Nasledujúci graf 1 znázorňuje počet respondentov a výšku ich mesačných výdavkov na nákup mliečnych výrobkov zatriedených podľa výšky voľných disponibilných mesačných prostriedkov na osobu v domácnosti v Sk.



Graf. 1 Výška mesačných výdavkov na mliečne výrobky podľa výšky disponibilných mesačných prostriedkov na osobu v domácnosti (v Sk)
Prameň: vlastný výskum

Pri zisťovaní závislosti medzi intenzitou vplyvu jednotlivých vyselektovaných nástrojov marketingovej komunikácie a výškou mesačných výdavkov na mliečne výrobky sme dospeli k výsledkom, ktoré sú prezentované v nasledujúcich tabuľkách a grafoch.

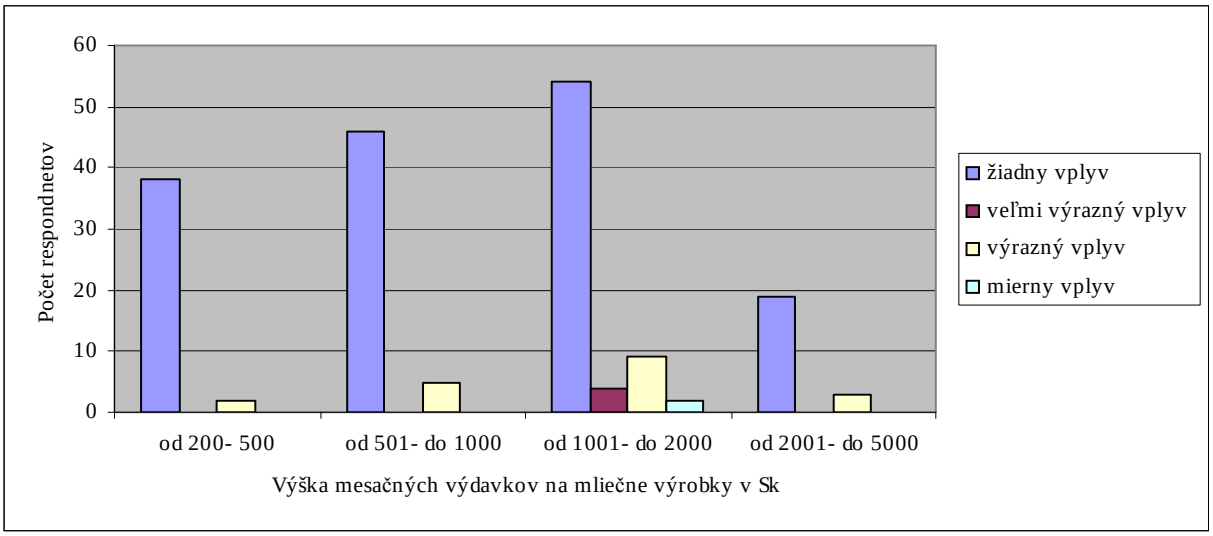
Nasledujúcim testovaním sme sledovali závislosť dvoch premenných, a to výškou mesačných výdavkov na mliečne výrobky a intenzitou reklamy v časopisoch.

Tabuľka 2 Závislosť medzi výškou mesačných výdavkov na mliečne výrobky a intenzitou vplyvu reklamy v časopisoch

Intenzita vplyvu reklamy v časopisoch	Výška mesačných výdavkov na mliečne výrobky v Sk				
	od 200- 500	od 501- do 1000	od 1001- do 2000	od 2001- do 5000	Celkový súčet
žiadny vplyv	38	46	54	19	157
veľmi výrazný vplyv	-	-	4	-	4
výrazný vplyv	2	5	9	3	19
mierny vplyv	-	-	2	-	2
celkový súčet	40	51	69	22	182

Prameň: vlastný výskum

V prípade testovania závislosti výškou mesačných výdavkov na mliečne výrobky a intenzitou vplyvu reklamy v časopisoch, je chí - kvadrát testovacia štatistika 12, 6, kritická hodnota je 16,92. Keďže hodnota chí - kvadrát je v tomto prípade nižšia než kritická hodnota, medzi premennými nie je závislosť, sú navzájom nezávislé. Výsledok odporuje doteraz získaným výsledkom z iných nami realizovaných výskumov, podľa ktorých by mal byť vplyv tohto nástroja marketingovej komunikácie natoľko významný, že ovplyvní spotrebiteľa pri nákupe. Podľa doterajších zistení by malo platiť, že čím je intenzita vplyvu reklamy v časopisoch vyššia, mali by byť aj výdavky na mliečne výrobky vyššie. Výsledky uvedeného testovania to nepotvrdzujú.



Graf 2 Výška mesačných výdavkov na mliečne výrobky podľa vplyvu reklamy v časopisoch
Prameň: vlastný výskum

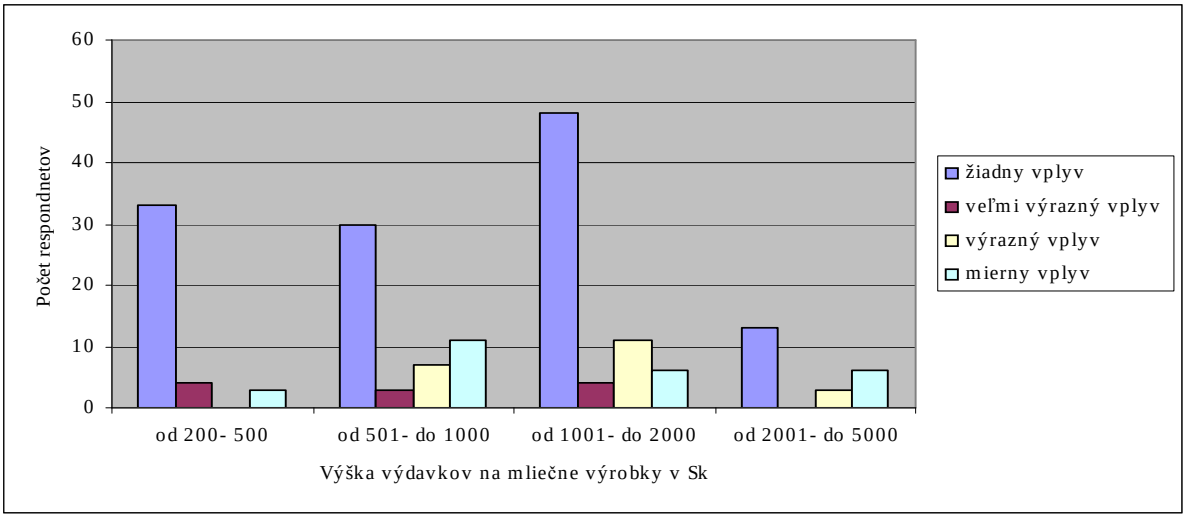
Ďalším testovaním sme zisťovali, či existuje závislosť medzi výškou mesačných výdavkov na mliečne výrobky a intenzitou vplyvu reklamy v televízii.

Tabuľka 3 Závislosť medzi výškou mesačných výdavkov na mliečne výrobky a intenzitou vplyvu reklamy v televízii

Intenzita vplyvu reklamy v televízii	Výška mesačných výdavkov na mliečne výrobky v Sk				
	od 200- 500	od 501- do 1000	od 1001- do 2000	od 2001- do 5000	Celkový súčet
žiadny vplyv	33	30	48	13	124
veľmi výrazný vplyv	4	3	4	-	11
výrazný vplyv	-	7	11	3	21
mierny vplyv	3	11	6	6	26
celkový súčet	40	51	69	22	182

Prameň: vlastný výskum

V prípade testovania závislosti výškou mesačných výdavkov na mliečne výrobky a intenzitou vplyvu reklamy v televízii, je chí - kvadrát testovacia štatistika 17,87, kritická hodnota je 16,91 a miera tesnosti závislosti je 0,29. Z testovania vyplýva, že medzi sledovanými premennými existuje silná závislosť. Na základe testovania možno konštatovať, že čím je intenzita vplyvu reklamy v televízii vyššia, tým sú vyššie aj mesačné výdavky na mliečne výrobky.



Graf 3 Výška mesačných výdavkov na mliečne výrobky podľa vplyvu reklamy v televízii
Prameň: vlastný výskum

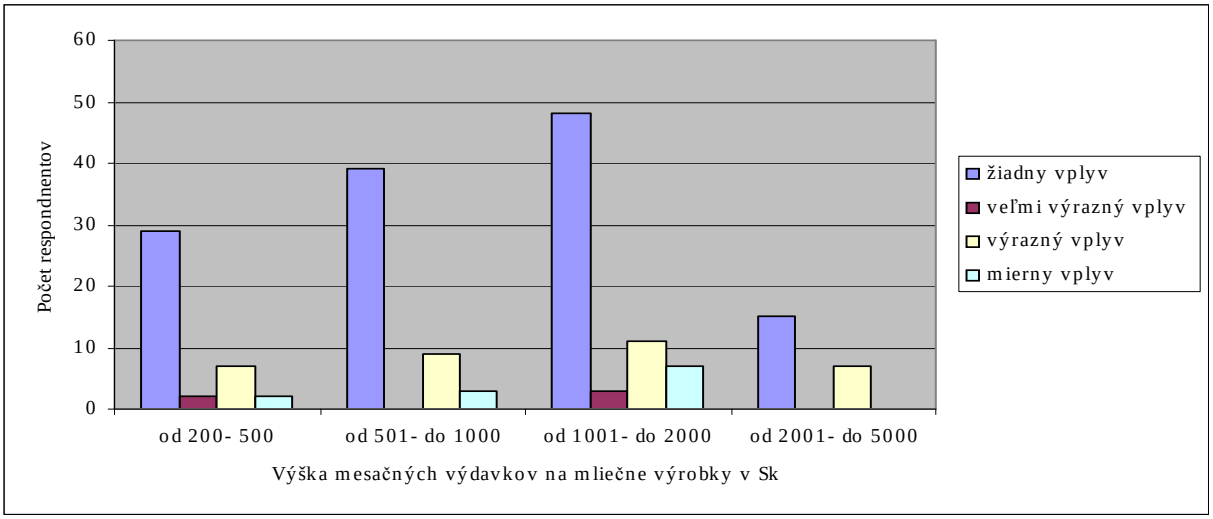
Nasledujúce testovanie sledovalo existenciu závislosti medzi výškou mesačných výdavkov na mliečne výrobky a intenzitou vplyvu reklamy v letákoch a plagátoch.

Tabuľka 4 Závislosť medzi výškou mesačných výdavkov na mliečne výrobky a intenzitou vplyvu reklamy na letákoch a plagátoch

Intenzita vplyvu reklamy v letákoch a plagátoch	Výška mesačných výdavkov na mliečne výrobky v Sk				
	od 200-500	od 501- do 1000	od 1001- do 2000	od 2001- do 5000	Celkový súčet
žiadny vplyv	29	39	48	15	131
veľmi výrazný vplyv	2	-	3	-	5
výrazný vplyv	7	9	11	7	34
mierny vplyv	2	3	7	-	12
celkový súčet	40	51	69	22	182

Prameň: vlastný výskum

V prípade zisťovania závislosti medzi výškou mesačných výdavkov na mliečne výrobky a intenzitou vplyvu reklamy v letákoch a plagátoch sme dospeli k nasledujúcemu výsledku. Keďže testovacia štatistika bola v tomto prípade 8,96, a kritická hodnota 16,91, tieto dve premenné sú nezávislé.



Graf 4 Výška mesačných výdavkov na mliečne výrobky podľa vplyvu reklamy v letákoch a plagátoch
Prameň: vlastný výskum

Záver

V prípade prvého testovania sme dospeli k záveru, že čím je výška mesačných voľných disponibilných prostriedkov na osobu v domácnosti vyššia, tým sú vyššie aj mesačné výdavky na mliečne výrobky. V prípade testovaní závislosti medzi výškou mesačných výdavkov na mliečne výrobky a vplyvom reklamy v televízii sa nám potvrdila silná závislosť. Pri rovnakom testovaní reklamy v časopisoch, reklamy v letákoch a plagátoch, jednotlivo, sa nám potvrdila nezávislosť sledovaných premenných. Bolo to prekvapujúce zistenie, pretože uvedené nástroje marketingovej komunikácie boli nástrojmi, ktoré v rámci už nami realizovaných výskumov, respondenti v dotazníkoch označovali ako nástroje, ktoré ich najviac ovplyvňujú pri rozhodovaní pre nákup mliečnych výrobkov. Na základe výstupov, už doteraz uskutočnených výskumov, sa prikláňame k tým výsledkom, ktoré dokazujú význam vplyvu skúmaných nástrojov marketingovej komunikácie na spotrebiteľské správanie vybranej skupiny spotrebiteľov pri mlieku a mliečnych výrobkoch. Z uvedeného vyplýva, že pre podniky pôsobiace v odvetví mliečnych výrobkov má význam investovať do uvedených nástrojov marketingovej komunikácie. Podľa našich výsledkov, sú tieto nástroje marketingovej komunikácie pri mliečnych výrobkoch viac alebo menej účinné a vložené finančné prostriedky podnikov na tento účel by mali priniesť minimálne nárast objemu nákupov zo strany spotrebiteľov. Získané výstupy zároveň môžu slúžiť ako cenný zdroj informácií a východisko pre plánovanie a realizáciu ďalších aktivít v spracovateľských podnikoch pôsobiacich na trhu mlieka a mliečnych výrobkov. V tomto zmysle je nevyhnutné, aby podniky rešpektovali nielen požiadavky spotrebiteľov vo vzťahu k ich produktom, ale zároveň aj sledovali, ktoré z nástrojov marketingovej komunikácie ovplyvnia nákupné rozhodovanie spotrebiteľa. V prípade, že podniky poznajú svojich spotrebiteľov, ich požiadavky a nástroje marketingovej komunikácie, na ktoré sú respondenti citliví, môže podnik pristúpiť k ďalším krokom pri svojom rozhodovaní vo vzťahu k trhu.

Kontakt:
Ing. Hroncová Ivana
Ústav vedy a výskumu UMB
Cesta na amfiteáter 1
974 01 Banská Bystrica
t. č. 048/ 446 6234, ivana.hroncova@umb.sk

Zvyšovanie konkurencieschopnosti slovenského vysokého školstva

Ľubica Hurbánková

Abstract

In this paper I introduce possibilities to increase of slovak higher education competitiveness. Paper is divided into two parts. In the first of them I attend variations in the higher education market, autonomy and competition as basic elements of academical policy. In the second part I analyse particular regions of competition and final conceptions.

Úvod

Zvyšovanie konkurencieschopnosti vysokého školstva je jedným z cieľov deklarovaných tzv. bolonskou deklaráciou. Vytvorenie inštitúcií schopných života a konkurencie by malo byť dosiahnuté ďalším rozvojom existujúcich inštitúcií. Schopnosť konkurencie a vysoká úroveň sú predpoklady, za ktorých sa môže vysoká škola profilovať a uplatniť.

1. Zmeny na trhu vysokého školstva, autonómia a konkurencia – základné prvky vysokoškolskej politiky

V poslednom období väčšina vysokých škôl rozširuje svoju ponuku a uprednostňuje široké postavenie na trhu pred výrazným vymedzením. Existuje ale aj opačná tendencia niektorých vysokých škôl, najmä univerzít s veľkou tradíciou, ktoré sa zameriavajú na svoju pôvodnú základnú koncepciu a ich konkurenčná stratégia spočíva práve v jednoznačnom vymedzení.

Známe sú tieto prvky vzájomnej konkurencie medzi vysokými školami:

- *učitelia* – prostriedkom konkurencie v tejto oblasti sú predovšetkým:
  a) pracovné podmienky a možnosti tvorivého uplatnenia,
  b) získavanie talentovaných učiteľských síl z praxe,
  c) eliminovanie formálnych prekážok pri získavaní učiteľov pri zachovávaní požiadaviek vysokej kvality,
- *študenti* - v tejto oblasti je najvýznamnejším nástrojom konkurencie medzi vysokými školami:
  a) miera zamerania na záujmy študujúcich,
  b) kvalita a rozsah individuálnej starostlivosti o študentov,
  c) ohľad na možnosti uplatnenia absolventov daného štúdia v praxi,
- *postavenie v sieti vysokých škôl* – ide o problematiku vzťahu k iným vysokým školám. Z tohto hľadiska sú dôležitými nástrojmi konkurencie:
  a) celková profilácia vysokej školy,
  b) zameranie študijných programov a usporiadanie výuky,
  c) spolupráca s inými vysokými školami,
- *partneri v praxi* – úspešnosť vysokej školy v konkurencii o partnerov v praxi predovšetkým závisí na:
  a) rozsahu, intenzite, kvalite a trvalosti kooperatívnych vzťahov s partnermi v hospodárskej a ďalších sférach praxe (štátna správa, kultúra a pod.) v otázkach obsahových a personálnych (prechod pracovníkov z praxe na vysokú školu a pod.),
  b) úzkej spolupráci vo výskume (aplikovaný výskum),
  c) miere spolupráce pri ďalšom vzdelávaní.

Zvyšovaním konkurencie medzi vysokými školami sa vysokoškolská politika približuje svojimi cieľmi a poňatím fungovania hospodárskej sfére. Presadzuje sa snaha úspešného presadenia sa na trhu. Z toho vyplýva, že jedným z výrazných trendov, presadzujúcich sa na trhu terciárneho vzdelávania, je ekonomizácia jeho inštitúcií.

Ekonomizácia inštitúcií terciárneho vzdelávania

Na ekonomizáciu vedy, resp. vysokého školstva sa pozerá predovšetkým inštitucionálne. Nemala by sa však zanedbávať súvislosť medzi ekonomizáciou organizačných foriem a ekonomizáciou obsahu, resp. účelovou orientáciou.

Ekonomizácia formy

Sem patrí nielen premena vysokej školy na podnik, ale aj premena študentov na zákazníkov. Predpoklad, že sa tým výuková ponuka automaticky zameria na potreby študujúcich, odráža iba časť pravdy. Ak sú študenti zákazníkmi, je výsledkom dobre pripravená prednáška a nároční poslucháči, ktorým leží na srdci kvalita výuky. Prvým dôsledkom toho je, že vyučujúci pri zle pripravenej výuke sú študentmi zle hodnotení. Druhým dôsledkom toho je, že zákazníci platia. Kto nemôže, či nechce platiť, nie je zákazník. A kto platí, chce za to zodpovedajúcu protislužbu. Preto je premena študentov na zákazníkov nutne spojená so zavedením poplatkov. Tu sa ekonomizácia stáva problémom sociálnej starostlivosti, spravodlivého rozdeľovania, rovnosti šancí.

Presadzovaná prax priamej podpory výskumných projektov obmedzuje konkurenciu a pripomína riadenie výskumu. Vysokoškolská efektívnosť nie je výsledkom riadenia a s tým súvisiacich hierarchických štruktúr. Podniková efektívnosť sa vyjadruje v input-output reláciách a v príslušných ziskoch. V oblasti vysokého školstva však efektívnosť nemožno chápať ako výsledok input-output relácií. Vysoké školy nie sú organizácie pre realizáciu dopredu definovaných cieľov efektívnosti, kvalita vysokých škôl sa neprejavuje v podriaďovaniu sa akýmkoľvek požiadavkám zvonku. Ukazuje sa v konkurencii, do ktorej môžu vysoké školy vstupovať iba za predpokladu, že majú rozsiahlu autonómiu.

Ekonomizácia obsahu

Ľudia celoživotne investujú do svojho ľudského kapitálu, aby udržali, resp. nadobudli svoju pracovnú schopnosť a schopnosť konkurencie na trhu práce. Tomuto trendu sa obsah výuky musí nutne prispôsobovať.

Konkurencia na trhu terciárneho vzdelávania

Vysoké školy v súčasnosti pociťujú stupňujúcu obmedzenosť zdrojov. Petície, protesty, demonštrácie však nepomôžu. Pre spoločnosť sa stáva prospešným, keď konkurencia núti aktérov, aby boli lepší ako iní, keď ich za to odmenia. Ak sa pozeráme na vysoké školstvo z tejto perspektívy, zistíme, že konkurencia je buď systematicky eliminovaná, alebo že na nevhodné stimulačné systémy doplatia šikovní a zdatní študenti.

V hospodárskej súťaži má výrobca vyššie zisky, ak je po jeho výrobkoch väčší dopyt, ako má jeho konkurent. V konkurencii vysokých škôl sú učitelia naopak odsudzovaní, ak sa na výuku lepšie pripravujú ako ich konkurenti. Ich zaťaženie skúškami stúpa neúmerne, majú menej času na výskum, na budovanie vlastnej kariéry, čo rozhoduje o ich budúcich príjmoch.

Vysoká škola sa stáva elitným tímom, že konkuruje o najlepších študentov. Semináre a prednášky s nadanými študentmi, ktorí prejavujú záujem o študijný odbor, stimulujú výskum a výuku, a naopak s pasívnymi a nemotivovanými študentmi unavujú. Najlepší študenti sa získavajú dobrými študijnými programami a dobrými vyučujúcimi. Čo je dobrý študijný program môže rozhodnúť iba konkurencia, v jej rámci sa cestou pokusu a omylu presadí to najlepšie. Snaha po porovnateľnosti cestou zjednocovania, harmonizácie a pod. študijných programov a ich začleňovanie do dopredu daných bakalárskych či magisterských modelov v podstate odporuje myšlienke konkurencie v oblasti vysokého školstva.

Študenti volia, a to stále viac v celosvetovom meradle, miesto svojho vzdelávania na základe kvality a možností uplatnenia. Pre vysoké školy, ktoré by spĺňali cieľ čo najväčšej

schopnosti rozvoja a konkurencie, musí vysokoškolská politika vytvoriť potrebné rámcové podmienky.

Jednotlivé reformné zmeny by nemali byť plnené izolovane, ale z existujúcich dielčích námetov a zámerov by mal byť vytvorený celkový koncept modernizácie vysokého školstva, zahŕňajúci aj cieľ zvýšenej konkurencie a atraktívnosti inštitúcií terciárneho vzdelávania.

V súvislosti s konkurenciou sa vynára otázka podnikovej orientácie vysokých škôl, čo vyúsťuje do dôležitej východiskovej tézy a kategorickej požiadavky dôsledného uplatnenia už spomínaného princípu autonómie vysokých škôl. Podniková orientácia vysokých škôl zlepšuje i postavenie učiteľa – bádateľa, zvyšuje inštitucionálnu silu a vplyv jednotlivých vzdelávacích zariadení terciárnej sféry.

Autonómia inštitúcií terciárneho vzdelávania

V praktickej vysokoškolskej politike sa zo spomenutých dôvodov požaduje rozsiahla autonómia pre vysoké školy, vrátane práva samostatného výberu študentov a vyberania poplatkov v rámci financovania štúdia.

Požiadavka dôsledného uplatnenia autonómie inštitúcií terciárneho vzdelávania priamo vyplýva z logiky konkurenčných systémov. Autonómia vysokých škôl je predpokladom ďalších podmienok pre konkurenciu medzi vysokými školami, t.j.

– decentralizovaná organizácia v záujme individuálnych, resp. lokálnych preferencií,
– rozhodovanie o optimálnom nasadení obmedzených zdrojov,
– delegovanie inovačnej zodpovednosti na jednotlivé vysoké školy.

Konkurencia požaduje dôkladne zladený systémový rámec. V tomto zmysle znamená uplatnenie princípu konkurencie vo vysokoškolskej politike predovšetkým vedomú rezignáciu na politické riadenie ponuky. Každá konkurencia musí mať svoje pravidlá hry, v oblasti vysokých škôl by sa tieto pravidlá mali týkať:

– podmienok prístupu na vysoké školy,
– uznávania absolutórií (diplomov),
– zabezpečovania transparentnosti a mobility,
– všeobecného celoštátneho zabezpečenia kvality štúdia.

Úlohou štátu v tomto kontexte je dbať o dodržiavanie spoločne vypracovaných pravidiel a zásad týkajúcich sa predpokladov na prijatie na vysokoškolské štúdium, o uznávaní jeho absolvovania, transparencie, mobility a čiastočne i kvality všetkých vysokých škôl a vyrovnávať podmienky konkurencie.

2. Oblasti konkurencie a cieľové predstavy

Koncepcia konkurencieschopnej vysokej školy sa týka nasledujúcich oblastí konkurencie a spočíva na týchto cieľových predstavách dielčích reformných polí jednotlivých vysokých škôl:

Tvorba profilov

- *vedecký a výskumný profil* – zásadný význam pre úspech konkurenčne organizovaného trhu vzdelávania má vlastný špecifický vedecký profil, ktorým sa vysoká škola prezentuje voči konkurujúcim vysokým školám. Jeho základom je tematické a problémové ťažisko výskumu, riešenie zodpovedajúcich problémov praxe. Cielenou podporou týchto ťažísk si vytvára vysoká škola predpoklady pre vznik Centers of Excellence, ktoré sú ústredným prvkom príťažlivosti vysokej školy v národnom i medzinárodnom meradle. Od toho odvodené konkrétne požiadavky sa prenášajú do sféry výuky.
- *profil vo výuke* – špecifický vedecký profil vysokej školy sa odráža i vo výuke. Podmienkou úspechu v konkurencii s inými inštitúciami terciárneho vzdelávania sú pružne

uskutočňované obsahové a organizačné reformy učebných plánov vyvolané stále častejšími zmenami vo vede i na trhu práce.

Zaistenie a rozvíjanie konkurenčných pozícií sa uskutočňuje stupňovaním efektívnosti:

- učebných obsahov,
- foriem štúdia,
- zloženia a práce učiteľského zboru,
- fungovania ostatných sietí.

Pokiaľ ide o obsah, možno predpokladať, že bude stále ťažšie uplatniť sa na trhu práce úzkym poňatím ponúkaných štúdií. Študijné programy je nutné udržiavať stále „up to date“, zdroje využívať čo najracionálnejšie, preskúmať existujúcu spoluprácu z hľadiska ďalších prínosov pre vysokú školu a rozvíjať výhodné a vytvárať nové, najmä v aplikačnom výskume, využívaním potenciálu hosťujúcich učiteľov a siete absolventov. Vo vzťahu k formám štúdia priebežne zvyšovať racionalitu a efektívnosť ich štruktúry, predovšetkým posilňovať zložku individuálneho štúdia.

Ďalším prvkom ovplyvňujúcim postavenie vysokej školy v konkurencii je rozsah, kvalita a intenzita starostlivosti o študentov a účastníkov ďalšieho vzdelávania jednak v priebehu samotného štúdia, jednak v procese začleňovania absolventov do praxe.

Manažment kvality

- *Mamažment kvality ako faktor konkurencie* – okrem špecifického profilu potrebujú vysoké školy v záujme konkurencie manažment kvality, systém rozvíjania a zaisťovania kvality v štúdiu i výuke, výskume i poskytovaní služieb. Vysoká škola, ktorá by svoju výlučnosť v tejto oblasti nemohla dokázať, v konkurencii neobstojí.
- *Mamažment kvality na autonómnej vysokej škole* – úprava vzťahov medzi štátom a autonómnou vysokou školou vyžaduje zvýšenú mieru transparentnosti v používaní prostriedkov a povinnosť skladať účty obsahujúce konkrétne zaobchádzanie s rozpočtom a zahŕňajúce všetky oblasti činnosti vysokej školy. Excelentnosť vysokej školy musí byť dokumentovaná a pri trvale neuspokojivých výsledkoch musí viesť k dôsledkom i pokiaľ ide o základné financovanie zo strany štátu.

  Opatrenia na zaistenie a zvyšovanie kvality v štúdiu a výuke musia presahovať bežné hodnotenie. Hodnotenie výuky vedie k nosným výsledkom až vtedy, keď je súčasťou celého procesu, ku ktorému patrí:
  - výber študujúcich samotnou vysokou školou,
  - akreditácia študijných programov a inštitúcií,
  - dlhodobé strategické plánovanie,
  - rozdeľovanie prostriedkov,
  - vývoj organizácie vnútri danej inštitúcie.
- *Mamažment kvality v správe vysokej školy* – predovšetkým poskytovanie služieb – starostlivosti o študentov a poradenstva, musí byť podrobené stálej kontrole. Spokojnosť študujúcich je rastúcou mierou podmienená kvalitou starostlivosti o nich. Začínajúc fázou uchádzania sa o štúdium a jeho zahájenia až po fázu ukončenia štúdia a vstupu do výkonu povolania. Udržovanie kontaktov s absolventmi a starostlivosť o ne má prvoradý význam.

Výber študujúcich

- *Zabezpečovanie kvality a tvorba profilov* – zvyšovanie a zabezpečovanie kvality štúdia a výuky začína výberom študujúcich. Ak je konkurencia myslená vážne, musí pôsobiť už pri vstupe na vysokú školu. Uchádzači o štúdium si vyberajú vysokú školu predovšetkým na základe jej profilu a vysoká škola si vyberá tých, ktorí sa zdajú byť vhodní. Toto je dôležitý predpoklad pre vlastný profil, pretože vytvorené profily možno trvalo etablovať

a využívať ako konkurenčnú výhodu len vtedy, ak sú študujúci schopní splniť príslušné požiadavky vysokej školy.

Financovanie

- *Tri piliere financovania* – základné financovanie vysokých škôl je úlohou štátu. V záujme určitej plánovacej istoty by toto financovanie malo prebiehať na podklade strednodobých dohôd s jednotlivými vysokými školami. Okrem toho musia mať vysoké školy neobmedzené právo samostatného získavania a vynakladania prostriedkov na splnenie svojich cieľov z iných zdrojov. Zásadne by malo financovanie vysokého školstva spočívať v troch pilieroch – základnom financovaní štátom, získavaní prostriedkov z mimorozpočtových zdrojov, príjmov z vlastného kapitálového kmeňa.
- *Optimálna organizačná forma* – otázkou je, aká právna forma najlepšie zaručuje vlastnú zodpovednosť v rozhodovaní vysokej školy o finančných záležitostiach. Najmä v otázkach odmeňovania to je významné. Vysoké školy by mali mať možnosť disponovať s nehnuteľnosťami, ktoré na ne boli prevedené na kapitálovom trhu za účelom získania prostriedkov jednak na krátkodobé zámery, jednak na investície dlhodobejšej povahy.
- *Príjmy vysokých škôl z mimorozpočtových zdrojov* – vlastné prostriedky môžu vysoké školy získať takto:
  - príjmy zo súkromných zdrojov (nadácií, darov, dedičstva a pod.),
  - trhové zhodnotenie výsledkov vlastného výskumu,
  - príjmy z aktivít ďalšieho vzdelávania,
  - príjmy z príspevkov študujúcich na krytie nákladov štúdia.

  Najmä na pozadí rastúceho významu celoživotného vzdelávania možno počítať zo strany podnikov so vzrastajúcou potrebou trhových ponúk. Tu môžu vysoké školy (nielen jednotliví učitelia ako vedľajšiu činnosť) v spolupráci s podnikmi vyvinúť a ponúknuť trhovo orientované produkty za trhové ceny.

  Finančné zhodnotenie výsledkov vlastného výskumu do značnej miery závisí od zákonnej úpravy autorských práv. Vysoká škola ako taká by mala mať uznaný nárok na podiel z autorského honorára voči svojim zamestnancom. To isté platí pre oblasť výskumnej spolupráce s podnikmi.

  Základné financovanie vysokých škôl a podpora výskumu, by mali byť orientované výkonovo.

Personál

- *Personál ako faktor konkurencie* – bez vynikajúcich pracovníkov a spolupracovníkov nemôže vysoká škola patriť k najlepším vedeckým a vzdelávacím inštitúciám. To platí pre každú jednotlivú oblasť, tieto však musia v koordinovanej súhre zaručovať vzájomnú podporu a dopĺňanie. Nevyhnutné je preto úzke vzájomné prepojenie jednotlivých oblastí činnosti vysokej školy.
- *Vysoká škola ako zamestnávateľ* – sem patrí právo vyberať pracovníkov (i z medzinárodne vypísaných konkurzov) a uzatvárať pracovné zmluvy bez ohľadu na to, či ide o vedecký, pedagogický, alebo iný personál. Nutná je možnosť slobodnej výmeny personálu medzi vysokou školou a podnikmi.
- *Trhovo primerané odmeňovacie štruktúry* – vysoká škola musí mať možnosť vyplácať svojim spolupracovníkom platy podľa výkonu a v súlade s trhovými podmienkami. Pre úspešnú personálnu politiku vysokej školy je dôležité:
  - nábor pracovníkov v medzinárodnom rozsahu,
  - odmeňovane a motivačné systémy orientované na výkon a úspech.

  Účelné je rozdelenie platov na časť pevnú a časť, ktorá závisí od výkonu jednotlivca.

- *Personálny rozvoj* – popri presvedčivému systému odmeňovania patrí k úspešnej personálnej politike vysokej školy aj koncepcia personálneho rozvoja, ktorá sa zameriava jednak na potreby vysokej školy, jednak na schopnosti jednotlivého spolupracovníka. Podobne ako v hospodárskej oblasti, musia vysoké školy rozvíjať i potenciál svojich pracovníkov pomocou cielených aktivít ďalšieho vzdelávania.

Otvorenosť voči trhu práce

- *Orientácia na trh práce ako faktor konkurencie* – dôraz na kompetenciu a kvalifikáciu nevyhnutnú pre trh práce, je vo výuke samozrejmou úlohou. Dôležitá je pritom sústavná spolupráca s hospodárskou sférou, ktorá musí vysokým školám tieto požiadavky artikulovať. Otvorenosť voči potrebám trhu práce ovplyvňuje tvorbu profilu vysokej školy, ktorá nemôže, ani to nie je jej úloha, reagovať na krátkodobé výkyvy.
- *Nové štruktúry štúdia* – orientácia na trh práce podporuje vytváranie nových štruktúr štúdia a študijných ponúk. Medzinárodne kompatibilné absolutóriá majú veľký význam pre zvýšenie profesijných predpokladov absolventov. Možnosť bakalárskych a magisterských absolutórií by mali vysoké školy využiť k zmenám študijných štruktúr a na integráciu profesijne kvalifikujúcich povinných vyučujúcich obsahov do študijných programov.

  Predpokladom akceptovania nových absolutórií na trhu práce je výuka, ktorá je preukázaná a sprehľadnená akreditačným riadením uskutočneným za účasti profesijných praktikov. Podniková sféra môže aktívne prispieť k úspechu nových študijných ponúk a absolutórií tým, že sa k nim hlási, že využíva možnosť účasti na tvorbe koncepcie študijných programov a že absolventov zamestná primerane ich kvalifikácií.

Internationalizácia

Internacionalizácia zahŕňa rozvíjanie medzinárodnej mobility študujúcich, vyučujúcich a ostatných pracovníkov, ako i medzinárodné zameranie študijnej ponuky. Veda síce vždy prekračovala hranice, ale v súčasnosti sa zvyšuje rýchlosť medzinárodnej mobility osôb, údajov, poznatkov a vedenia. Medzinárodná kompatibilita absolutórií musí byť zaistená i preto, aby prišlo k zvýšeniu počtu zahraničných študentov.

Transfer vedenia

Vysoké školy spolupracujú s hospodárstvom a zdokonaľujú transfer vedenia v rôznych formách.

- *Transfer vedenia – dôležitý faktor kvality a konkurencie* – transfer vedenia a technológií medzi vysokými školami a podnikmi má pre hospodársky rast, spoločenský a vedecký rozvoj značný význam a úžitok. Vysokým školám sa otvárajú nové témy a prístupy k oblastiam rýchleho technologického rozvoja, podniky získajú priamy prístup k „vedeckej obci“ a k know how v oblasti vzdelávania a výskumu.
- *Formy spolupráce* – zvláštny význam pre výmenu a vzájomné zúžitkovanie vedenia a praktických skúseností má:
  - výmena personálu na prechodnú dobu,
  - trhové využitie vynálezov a patentov vysokými školami,
  - poradenstvo zamerané na vysokoškolských učiteľov a absolventov týkajúce sa možností rozvíjania transferu vedenia (v spolupráci s podnikmi je v popredí napr. otázka podnikateľského osamostatnenia sa).

  Pri pohľade na výuku je spolupráca medzi vysokými školami a podnikmi prostriedkom na zvýšenie kvality v dvoch smeroch: na základe obsahovej a personálnej výmeny medzi vysokou školou a hospodárstvom môže študijná ponuka priebežne reflektovať meniace sa aktuálne požiadavky a potreby trhu práce.

Manažment / Správa / Administratíva

Vysoké školy koncipujú samy stratégiu vlastného vývoja a riadia tento vývojový proces. K tomu sú nevyhnutné efektívne riadiace štruktúry s presne vymedzenými právomocami. Okrem toho je nutná efektívna, na poskytovanie dokonalého servisu zameraná administratíva (správa).

Významným faktorom zvyšovania efektívnosti a konkurencieschopnosti autonómnej vysokej školy sú jej riadiace a správne štruktúry. Základnými prvkami sú jasne formulované a navzájom jednoznačne vymedzené právomoci v oblasti dohľadu, operatívneho riadenia a realizácie (prevádzania). Usporiadanie riadiacich a kontrolných štruktúr je nutné ponechať samotným vysokým školám.

Ako prvky koordinácie a riadenia na centrálnej i na nižších úrovniach sa na zahraničných vysokých školách osvedčujú tzv. dohody o cieľoch (výkonoch). Dohody tohto druhu sú účinným nástrojom strategického usmerňovania vysokých škôl zo strany štátu v zmysle celoštátnej vysokoškolskej politiky.

Záver

Pozíciu vysokej školy v konkurencii možno zaistiť a posilniť uplatňovaním týchto zásad:

– zreteľná diferenciácia voči iným vysokým školám a výskumným zariadeniam,
– rozvíjanie vlastných silných stránok a nezanedbávanie slabých, t.j. ponuku na základe skúseností udržiavať „blízko trhu“,
– vyhýbať sa honbe za „vzormi“, pretože by sa tým strácala jedna z konkurenčných výhod,
– výskum, pretože je finančne náročný, uskutočňovať v spolupráci s inými,
– intenzívne pracovať na kvalite vlastnej „značky“.

Vysoké školy budúcnosti by mali byť organizované tak, aby rovnako pružne ako podniky, s ktorými spolupracujú, sa mohli adaptovať na zmeny na trhu a tieto zmeny ovplyvňovať. K tomu musí mať vysoká škola čo najväčšiu autonómiu pri utváraní svojho profilu, pri hospodárení so svojimi prostriedkami, výbere personálu a usporiadaní svojej organizácie.

Zoznam použitej literatúry

[1] Roth, O.: Zvyšování konkurenceschopnosti evropských vysokých škol; rozbor prvků ovlivňujících konkurenceschopnost. In: Academia 1/2005
[2] Universität am Scheideweg : Herausforderungen – Probleme – Strategien. Publikation der Akademischen Kommission der Universität Bern. Zuerich: Hochschulverlag AG an der ETH, 1998
[3] Wegweiser der Wissensgesellschaft, zur Zukunfts –und Wettbewerbsfähigkeit unserer Hochschulen. Hochschulrektorenkonferenz. Berlin: Bundesvereinigung der deutschen Arbeitgeberverbände, 2003

Autor

Ing. Ľubica Hurbánková
Katedra štatistiky
Fakulta hospodárskej informatiky
Ekonomická univerzita v Bratislave
Dolnozemská cesta 1
852 35 Bratislava
tel: 02/67 295 728
e-mail: lu.pod@post.sk

DIFERENČNÉ ROVNICE NECELOČÍSELNÉHO RÁDU A ICH RIEŠENIE POMOCOU SOFTVÉRU MATHEMATICA

Anton HUŤA

Abstract The aim of the paper is to introduce new form of the generalized difference equation of the fractional order. The exact solution is possible using software Mathematica.

Key words: difference equation of the fractional order, Mathematica

Základná myšlienka riešiť diferenciálne rovnice neceločíselného rádu pochádza zo sedemnásteho storočia. Špeciálnym prípadom bol prípad rádu ½, resp. *n*/2, pre $n \in$ N, kedy hovoríme o tzv. semidiferenciálnych rovniciach. Nasledujúca definícia diferenčnej rovnice neceločíselného rádu je nová a jej autorom je autor tohoto článku. Cieľom príspevku je ukázať na možnosť efektívneho riešenia takéhoto typu rovníc pomocou softvéru Mathematica.

Materiál a metódy

Lineárnu diferenčnú rovnicu neceločíselného rádu uvažovať v tvare (v zjednodušenom prípade)

$$\sum_{k=0}^{n} a_k \nabla^{\alpha_k} y(t) = b(t), \text{ pre } n \geq 0,$$

Využitím vzťahu

$$\nabla^{\alpha} = \sum_{k=0}^{\infty} (-1)^k \binom{\alpha}{k} E^{-k} \quad ,$$

čo predstavuje definíciu zovšeobecnej diferencie, dostávame

$$\sum_{j=0}^{\infty} (-1)^j E^{-j} \sum_{k=0}^{n} \binom{\alpha_k}{j} y(t) = b(t).$$

Posledný vzťah môžeme považovať za definíciu lineárnej diferenčnej rovnice neceločíselného rádu. Tento tvar diferenčnej rovnice neceločíselného rádu je v súlade s definíciou klasických diferenčných rovníc celočíselného rádu. Špeciálnym prípadom diferenčných rovníc neceločíselného rádu sú klasické diferenčné rovnice celočíselného rádu. Keďže α_k, k = 0, 1, ..., n môžu byť ľubovoľné reálne čísla, zahŕňa daný tvar aj prípady $0 < \alpha_k < 1$ známe z dostupnej literatúry.

Výsledky a diskusia

Na exaktné vyjadrenie riešenia diferenčnej rovnice neceločíselného rádu sa dá úspešne použiť softvér *Mathematica*. Popísaný postup umožňuje vyjadriť riešenie v exaktnom tvare pri ľubovoľne presnom priblížení, t.j. použitý rozvoj Taylorovho radu obsahuje dostatočný počet členov Taylorovho polynómu

$$\sum_{j=0}^{m} (-1)^j E^{-j} \sum_{k=0}^{n} \binom{\alpha_k}{j} y(t) = b(t).$$

Zápis v softvéri Mathematica

```
(* RSolve [rekurentný vzťah, y[n], n, počiatočné podmienky]
rieši rekurentnú rovnicu neceločíselného rádu
pre y[n] s danými začiatočnými podmienkami
a // Simplify zjednoduší, upraví výsledok*)
RSolve
[{y[t] - 2 y[t - 1] + y[t - 2] + Sum[((-1) ^j) Binomial[Sqrt[2], j] y[t - j],
{j, 0, 2}] + y[t] == 0, y[0] == 0, y[1] == 1 }, y[t], t] // Simplify
```

Riešenie

$$\left\{\left\{y[t] \to \frac{2^{-t/2}\, 3^{1-\frac{t}{2}} \left(4-\sqrt{2}\right)^{\frac{1+t}{2}} \left(2+\sqrt{2}\right) \text{Sin}\left[t\, \text{ArcTan}\left[\frac{\sqrt{2\left(9-5\sqrt{2}\right)}}{2+\sqrt{2}}\right]\right]}{\sqrt{22+5\sqrt{2}}}\right\}\right\}$$

Pri riešení diferenčnej rovnice v klasickom zmysle potrebujeme toľko začiatočných hodnôt, koľkého je rádu. Hoci softvér *Mathematica* umožňuje exaktné vyjadrenie riešenia získanej diferenčnej rovnice pri rastúcom počte členov Taylorovho polynómu, nadobúda vyjadrenie neúnosnú zložitosť, je výhodnejšie počítať s priblíženiami.

Zápis v softvéri Mathematica

```
(*vynásobením hodnotou 1. dostaneme numerické priblíženie *)
RSolve
[{y[t] - 2 y[t - 1] + y[t - 2] + 1. Sum[((-1) ^j) Binomial[Sqrt[2], j] y[t - j],
{j, 0, 2}] + y[t] == 0, y[0] == 0, y[1] == 1 }, y[t], t] // Simplify
```

Riešenie

```
{{y[t] → 0. 0.656479^t Cos[0.522048 t] + 3.05476 0.656479^t Sin[0.522048 t]}}
```

Pomocou softvéru Mathematica sa dá riešiť aj sústava diferenčných rovníc neceločíselného rádu.

Zápis v softvéri Mathematica

```
(* RSolve rieši sústavu diferenčných rovníc
 neceločíselného rádu s danými začiatočnými podmienkami *)
RSolve[{a[n + 1] - 3  Sum[((-1) ^j) Binomial[Sqrt[2], j] b[n - j], {j, 0, 1}] == 1,
        a[n + 1] + b[n] == n,
        a[0] == b[0] == 0}, {a[n], b[n]}, n] // Simplify
```

Riešenie

$$\left\{\left\{a[n] \to \frac{1}{3\left(4-3\sqrt{2}\right)^6}\left(9\left(2^{4-\frac{3n}{2}}\left(-577+408\sqrt{2}\right)\left(2^{3n/2}-3^n\right)-4\left(-4756+3363\sqrt{2}\right)n\right)+\right.\right.$$
$$\left.\left(-302568+213948\sqrt{2}+1259\, 2^{\frac{11}{2}-\frac{3n}{2}}\, 3^n-1187\, 2^{4-\frac{3n}{2}}\, 3^{1+n}-72\left(-3363+2378\sqrt{2}\right)n\right)\text{UnitStep}[-2+n]\right),$$
$$\left.b[n] \to \frac{2^{3-\frac{3n}{2}}\left(-17+12\sqrt{2}\right)\left(2^{3n/2}-3^n\right)+\left(280-198\sqrt{2}\right)n}{\left(4-3\sqrt{2}\right)^4}\right\}\right\}$$

Záver

Diferenčné rovnice a ich riešenie nachádza uplatnenie v mnohých nematematických disciplínach (deje v prírode, spoločnosti a inde ako v ekonómii, genetike, demografii, medicíne, fyzike, ekológii, elektrotechnike, sociológii a inde). Ide tu o popis zmien v správaní sa pozorovaných dejov a ich priebehov. V procese riešenia uvedených rovníc sa vyžaduje spoľahlivosť modelovania a predpovede. Vo všeobecnosti je snaha využívať modely spoľahlivé zo stanoviska istých kritérií, pretože niektoré modely nesmia byť prijateľné. S tými súvisí výber zodpovedajúceho matematického modelu popisujúceho chovanie sa skúmaného procesu. Spoľahlivý model nám poskytuje možnosť použiť matematický aparát umožňujúci získať matematické záverečné rozhodnutia. Zavedenie diferenčných rovníc neceločíselného rádu rozširuje možnosť konštrukcie adekvátnych modelov reálneho sveta.

Abstrakt V príspevku je definovaný nový tvar zovšeobecnej diferenčnej rovnice neceločíselného rádu. Exaktné vyjadrenie jej riešenia je možné pomocou softvéru Mathematica.

Kľúčové slová: diferenčná rovnica neceločíselného rádu, Mathematica

Literatúra

[1] ORTIGUEIRA, M. D. : *Introduction to fractional linear systems*. Part 2: *Discrete-Time case*. Vol. 147, No. 1, s. 71 – 78, 2000.

[2] OSTALCZYK, P.W.: *The Time-varying Fractional Order Difference Equations*. ASME Designing Engineering Technical Conferences and Computers and Information in Engineering Conference. s. 1-9, 2003.

[3] PODLUBNY, I.: *Fractional Differential Equations Mathematics in science and engineering.* s. 313-335, 1999 ISBN 0-12-558840-2

[4] PODLUBNY, I.: *Derivácie neceločíselného rádu: História, teória, aplikácie.* Prednáška. Jasná, Konferencia Slovenskej matematickej spoločnosti, 2002

Kontaktná adresa

RNDr. Anton Huťa, PhD, KAGDM, Fakulta matematiky, fyziky a informatiky UK, Mlynská dolina, 842 48 Bratislava, telefón: 02 60295 255
e-mail: Anton.Huta@fmph.uniba.sk

VÝVOJ REÁLNYCH MIEZD V SR

Jozef Chajdiak[1]

V nedeľu 25. 9. 2005 v televíznej relácii „O 5 minút 12“ diskutovali o sociálno-ekonomickom vývoji predseda vlády SR M. Dzurinada a podľa septembrových preferencií budúci predseda vlády SR R. Fico. Predstaviteľ pravej strany politického spektra a predstaviteľ strany, ktorá zaplnila ľavú časť politického spektra. V rámci diskusie zaznel aj ostrý spor ohľadne vývoja reálnych miezd na Slovensku. Ako to už býva pri medializovaných politických sporoch, každý z oboch diskutujúcich mal svoju pravdu. Aká je to pravda?

Štatistický úrad zisťuje, okrem iného, aj výšku miezd. Výšky miezd na zamestnanca v národnom hospodárstve sa zisťujú na základe štvrťročného štatistického výkazníctva. Z neho sa určuje priemerná nominálna mesačná mzda zamestnanca v národnom hospodárstve v SR (v tisíc Sk) vo verzii za príslušný štvrťrok a vo verzii za celý rok a tiež index nominálnej mzdy a index reálnej mzdy. Oba indexy sa počítajú k rovnakému obdobiu minulého roka. Index reálnej mzdy je vypočítaný ako podiel indexu nominálnej mzdy a indexu spotrebiteľských cien.

V roku 2004 bola priemerná mesačná mzda zamestnanca v národnom hospodárstve 15825 Sk a v 2. štvrťroku 2005 bola 16737 Sk.

Vývoj priemerných mesačných miezd zamestnancov v národnom hospodárstve (MZDA) v rokoch 1994 až 2004 a zodpovedajúceho indexu reálnych miezd (IRM) je v nasledujúcej tabuľke č. 1 (Prameň: www.statistics.sk).

Tab. 1

ROK	MZDA	IRM	EUR0MZDA
1994	6 294	1.032	166
1995	7 195	1.040	187
1996	8 154	1.071	212
1997	9 226	1.066	243
1998	10 003	1.027	253
1999	10 728	0.969	243
2000	11 430	0.951	268
2001	12 365	1.010	286
2002	13 511	1.058	316
2003	14 365	0.980	346
2004	15 825	1.025	395

Vidíme, že v období vlády p. Dzurindu až tri roky bol pokles reálnych miezd a v roku 2000 oproti roku 1999 skoro o 5 %.

Priemerný ročný index rastu reálnych miezd za roky 1999 až 2004 je 0.998, t. j. za vlád p. Dzurindu každy rok poklesla reálna mzda v priemere o 0.2 %. Pre zaujímavosť za obdobie 1994 až 1998 (časť vlád p. Mečiara, keď budeme abstrahovať od symbolického polroku vlády p. Moravčíka) bol priemerný ročný index rastu reálnych miezd 1.047, t.j. v rokoch 1994 až 1998 každoročne reálna mzda narástla o 4.7 %.

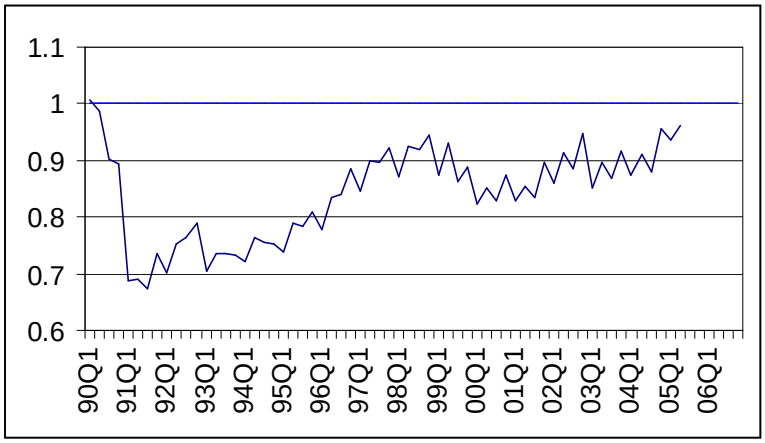
Keď pozrieme na vývoj v štvrťročnom pohľade je otázkou, či pre hodnotenie vývoja za obdobie vlád p. Dzurindu zobrať za základ 3. alebo 4. švrťrok 1998. Vývoj je však

[1] Doc. Ing. Jozef Chajdiak, CSc. – mimoriadny profesor na Katedre štatistiky Fakulte hospodárskej informatiky Ekonomickej univerzity v Bratislave

v oboch prípadoch dosť podobný. V kumulatívnom pohľade rast reálnych miezd bol v 4. štvrťroku 1998, v 1. a 2. štvrťroku 1999, potom od 3. štvrťroku 1999 po 4. štvrťrok 2004 nasledoval priemerný pokles reálnych miezd a v 1. a 2. štvrťroku 2005 sme sa opäť dostali do priemerného rastu. Oproti 3. štvrťroku 1998 k 2. štvrťroku 2005 priemerne štvrťročne vzrástla reálna mzda o +0.4 % (4. štvrťrok 2004 bo bol ešte pokles o 0.1 %, v 1. štvrťroku 2005 už vzrast o +0.1 %). V pohľade štvrťročných čísiel zlom vo vývoji nastal od 1. štvrťroka 2005.

Na vývoj miezd môžeme pozerať nielen z pohľadu špičkových politikov. Zaujímavé môže byť porovnanie ich vývoja s rokom 1989. Reálne máme vyššie mzdy ako za socializmu?. Iný pohľad predstavuje ich prepočet na eurá.

Na obr. 1 je znázornený vývoj reálnych miezd zamestnancov v národnom hospodárstve SR, pričom bázu (hodnota 1=100 %) predstavujú jednotlivé štvrťroky roku 1989. Z obrázku vidíme, že okrem prvého štvrťroku 1990 stále zarábame menej ako v roku 1989. V roku 1991 sme sa dostali na dno zhruba na úrovni dvoch tretín (0.673) zo zodpovedajúcej mzdy roku 1989. Vznik Slovenskej republiky opätovne spôsobil zreteľne viditeľný pokles reálnych miezd. Obdobie rokov 1994 až 1998 charakterizuje pomerne rovnomerný dynamický rast (až na úroveň 0.945 v štvrtom štvrťroku 1998). Za novej konštelácie moci v rokoch 1999 až 2002 vidíme priebeh v tvare písmena U (s dnom 0.829 v prvom štvrťroku 2001 a vrcholom 0.948 v štvrtom štvrťroku 2002). Súčasná pravicová vláda začala svoju vládu poklesom reálnych miezd (0.851 v prvom štvrťroku 2003) a ich postupným rastom (až na najlepšiu hodnotu po páde socializmu na úrovni 0.960 v druhom štvrťroku 2005 znamenajúcej, že stále v priemere zarábame o 4 % menej ako v roku 1989). Dúfame, že ostatný rast miezd vydrží čo najdlhšie a možno už za tejto vlády sa dostaneme nad hodnotu 1 a prekonáme úroveň socializmu.



Obr.1 Vývoj kumulatívneho indexu reálnych miezd (báza rok 1989)

Iný pohľad na naše mzdy dáva možnosť zavedenia eura k 1.1.2009. Priemerná mesačná mzda zamestnancov v národnom hospodárstve SR vyjadrená v eurách (do roku 1999 v ecu – XEU) je uvedená v tab. 1 v stĺpci EUROMZDA. Vidíme je postupný rast z 166 na 395 euro. V druhom štvrťroku 2005 je to už 430 euro. Keď si uvedomíme, že do 1.1.2009 máme už len 3 roky resp. 13 štvrťrokov, mzda na úrovni 600 euro nám dáva len utešujúci pocit, že v EU sú resp. budú aj krajiny, ktoré sú na tom ešte horšie.

Trendy posledného obdobia naznačujúce rast reálnych miezd a spevňovania kurzu SKK voči EUR by sme mali, napriek tvrdému politickému boju, zachovať a podľa možnosti aj prehĺbiť.

Kontakt na autora:
chajdiak@statis.biz
chajdiak@euba.sk

Comparison of Procrustes and Bookstein $2D$ coordinates in shape analysis: simulations, bagplots for landmarks and applications

Stanislav Katina

Abstract: In this paper we consider and compare Procrustes and Bookstein $2D$ coordinates in shape analysis on the base of simulations and graphs. Application of mentioned methods in biological anthropology is made using program routines implemented to R and S-PLUS statistical packages using statistical library from Katina (2004), Dryden and Mardia (1998), Rousseeuw and Ruts (1997) and new program routines.

1. Procrustes and Bookstein $2D$ coordinates

Let us consider n objects in $\mathcal{R}^2$ which are described by *configuration matrices* $\mathbf{X}_i = \left[\mathbf{x}_i^{(1)} \vdots \mathbf{x}_i^{(2)}\right] \in \left(\mathcal{R}^2\right)^k$, where $\mathbf{x}_i^{(m)} = \left(x_{i1}^{(m)}, ..., x_{ik}^{(m)}\right)^T$, $m = 1, 2$, $i = 1, 2, ..., n$ with respect to any orthogonal coordinate system. The *shape* $[\mathbf{X}]_S$ of a configuration $\mathbf{X} = \left[\mathbf{x}^{(1)} \vdots \mathbf{x}^{(2)}\right] \in \left(\mathcal{R}^2\right)^k$ is defined as the equivalent class of $\mathbf{X}$ with respect to the similarity transformations in $\mathcal{R}^2$, i.e. $[\mathbf{X}]_S = \left\{\mathbf{Y} = \left[\mathbf{y}^{(1)} \vdots \mathbf{y}^{(2)}\right] \in \left(\mathcal{R}^2\right)^k, \boldsymbol{\Gamma} \in SO(2), \mathbf{t} \in \mathcal{R}^2, a \in \mathcal{R}^+ : \mathbf{y}_j = a\boldsymbol{\Gamma}\mathbf{x}_j + \mathbf{t}, j = 1, 2, ..., k\right\}$, where $\boldsymbol{\Gamma}$ is rotation matrix, $SO(2)$ is the set of all rotations in $\mathcal{R}^2$, $\mathbf{x}_j^T$, resp. $\mathbf{y}_j^T$ are the rows of matrix $\mathbf{X}$, resp. $\mathbf{Y}$. We call $D = \left\{\mathbf{X} \in \left(\mathcal{R}^2\right)^k : \mathbf{x}_1 = \mathbf{x}_2 = ... = \mathbf{x}_k\right\}$ the diagonal of $\left(\mathcal{R}^2\right)^k$. Let $|\mathbf{x}|$ is the Euclidean norm of the vector $\mathbf{x} \in \mathcal{R}^2$, then $\|\mathbf{X}\| = \left(\sum_{j=1}^k |\mathbf{x}_j|\right)^{1/2}$ is the Euclidean norm of $\mathbf{X} \in \left(\mathcal{R}^2\right)^k$. The *centered configuration* $\mathbf{X}_c$ to $\mathbf{X} \in \left(\mathcal{R}^2\right)^k$ is $\mathbf{X}_c = \mathbf{X} - \mathbf{1}_k\overline{\mathbf{x}}^T$, where $\overline{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^k \mathbf{x}_j$ is the center of the points (landmarks) $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_k \in \mathcal{R}^2$ of matrix $\mathbf{X}$. Especially, $\mathbf{X} \in D$ if and only if $\|\mathbf{X}_c\| = 0$. The normed centered configuration $\mathbf{X}^*$ of a configuration $\mathbf{X} \in \left(\mathcal{R}^2\right)^k \setminus D$, $\mathbf{X}^* = \frac{\mathbf{X}_c}{\|\mathbf{X}_c\|}$, is called the *centered preshape* to $\mathbf{X}$. A *configuration* $\mathbf{X}$ is called *centered*, resp. *normed centered* if $\mathbf{X} = \mathbf{X}_c$, resp. $\mathbf{X} = \mathbf{X}^*$. We denote by $\widetilde{S}_2^k = \left\{\mathbf{X}^* : \mathbf{X} \in \left(\mathcal{R}^2\right)^k \setminus D\right\}$ the *preshape space*. If we reduce the definition of shape to the smaller equivalence class, we get $\widetilde{\mathbf{X}}^* = \{\mathbf{X}^* = \boldsymbol{\Gamma}\mathbf{X}^* : \boldsymbol{\Gamma} \in SO(2)\}$ for the shape of $\mathbf{X}$, where we esentially deal with normed centered configurations. By $\Sigma_2^k = \left\{\widetilde{\mathbf{X}}^* : \mathbf{X} \in \left(\mathcal{R}^2\right)^k \setminus D\right\}$ is denoted the *Kendall's shape space* (Ziezold, 1994).

We say that $\mathbf{X}_1, ..., \mathbf{X}_n$ are in *optimal position* or have *the best Procrustes fit* in the sense of "shape" if (Ziezold, 2003)

$$\begin{aligned}\sum_{1\le i\le j\le n} \left\|\mathbf{X}_i^{*\prime} - \mathbf{X}_j^{*\prime}\right\|^2 &= \inf_{\substack{\boldsymbol{\Gamma}_1,...,\boldsymbol{\Gamma}_n \in SO(2)\\ \mathbf{t}_1,...,\mathbf{t}_n \in \mathcal{R}^2}} \sum_{1\le i\le j\le n} \left\|a\boldsymbol{\Gamma}_i\mathbf{X}_i + \mathbf{1}_k\mathbf{t}_i^T - a\boldsymbol{\Gamma}_j\mathbf{X}_j - \mathbf{1}_k\mathbf{t}_j^T\right\|^2 \\ &= \inf_{\boldsymbol{\Gamma}_1,...,\boldsymbol{\Gamma}_n \in SO(2)} \sum_{1\le i\le j\le n} \left\|\boldsymbol{\Gamma}_i\mathbf{X}_i^* - \boldsymbol{\Gamma}_j\mathbf{X}_j^*\right\|.\end{aligned}$$

Then one can call the rows of $\mathbf{X}_i^{*\prime}$, $i = 1, ..., n$ as *Procrustes* $2D$ *coordinates*. Details of computations one can see in Goodall (1991).

Baseline size in $2D$ is the length between landmark 1 and 2

$$D_{12}(\mathbf{x}) = \|\mathbf{x}_2 - \mathbf{x}_1\| = \sqrt{\left(x_2^{(1)} - x_1^{(1)}\right)^2 + \left(x_2^{(2)} - x_1^{(2)}\right)^2},$$

where $\mathbf{x}_j = \left(x_j^{(1)}, x_j^{(2)}\right)^T, j = 1, 2$. *Bookstein* $2D$ *coordinates* $\mathbf{b}_j = \left(b_j^{(1)}, b_j^{(2)}\right)^T$, are the remaining coordinates of an object after translating, rotating and rescaling the baseline to $\mathbf{b}_1 = \left(b_1^{(1)}, b_1^{(2)}\right)^T = \left(-\frac{1}{2}, 0\right)^T$ and $\mathbf{b}_2 = \left(b_2^{(1)}, b_2^{(2)}\right)^T = \left(\frac{1}{2}, 0\right)^T$ so that

$$b_j^{(1)} = \frac{\left[\left(x_2^{(1)} - x_1^{(1)}\right)\left(x_j^{(1)} - x_1^{(1)}\right) + \left(x_2^{(2)} - x_1^{(2)}\right)\left(x_j^{(2)} - x_1^{(2)}\right)\right]}{D_{12}^2} - \frac{1}{2},$$
$$b_j^{(2)} = \frac{\left(x_2^{(1)} - x_1^{(1)}\right)\left(x_j^{(2)} - x_1^{(2)}\right) - \left(x_2^{(2)} - x_1^{(2)}\right)\left(x_j^{(1)} - x_1^{(1)}\right)}{D_{12}^2}.$$

So, we have *mapping* Ψ: $\mathbf{x}_j \rightarrow \mathrm{b}_j = \left(b_j^{(1)}, b_j^{(2)}\right)^T$ for $j = 1, ..., k$ and have to find the scale $c > 0$, the rotation $\mathbf{\Gamma}$, and the translation $\mathbf{t} = (t_1, t_2)^T$ such that $\Psi : \mathbf{x}_j \rightarrow \mathbf{b}_j = c\mathbf{\Gamma}(\mathbf{x}_j - \mathbf{t})$, $j = 3, 4, ..., k$ where $\mathbf{x}_j = \left(x_j^{(1)}, x_j^{(2)}\right)^T$ and $\mathbf{b}_j = \left(b_j^{(1)}, b_j^{(2)}\right)^T$ are coordinates of the j-th point before and after the transformation, $\mathbf{\Gamma}_{2\times 2}$ is rotation matrix $\mathbf{\Gamma} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$, where we rotate clockwise by θ radians. Finally, the mapping Ψ can be applied for all landmarks. To sum up, $\mathbf{B} = \left[\mathbf{b}^{(1)} : \mathbf{b}^{(2)}\right]$ is $k \times 2$ *matrix of Bookstein coordinates* (Dryden and Mardia, 1999).

Let we have $(k \times 2)$-configuration matrices $\mathbf{X}_i$, $i = 1, 2, ..., n$, then $\mathbf{B}_i$ are matrices of Bookstein coordinates derived from $\mathbf{X}_i$, $i = 1, 2, ..., n$.

2. Bivariate Tukey medians and bagplots for landmarks

Consider a bivariate data set, here a set of landmarks $X_j = \{\mathbf{x}_{1j}, ...\mathbf{x}_{nj}\}$, where $j = 1, ..., k$. The *halfplane location depth* of an arbitrary point $\theta_j \in \mathcal{R}^2$ relative to X_j is defined as

$$ldepth\left(\theta_j, X_j\right) = \min_{H_j} \#\left\{i : \mathbf{x}_{ij} \in H_j\right\}, i = 1, ..., n; j = 1, ..., k,$$

where H_j ranges over all closed halfplanes of which the boundary line passes through θ_j (Tukey, 1975). Obviously, a point θ_j outside the convex hull of X_j will have depth zero.

The *depth region of depth* l is defined as the set D_l of points θ_j with $ldepth\left(\theta_j, X_j\right) \geq l$. Equivalently, D_l is the intersection of all closed halfplanes that contains at least $n - l + 1$ observations, hence D_l is a bounded convex polytope, and $D_{l+1} \subset D_l$. The boundary of D_l is a convex polygon, which is called the *contour of depth* l (*l-th depth contour*, *l-th hull*). Therefore, each vertex of a depth contour is the intersection point of two lines, each passing through two observations. The *Tukey median* $\mathbf{t}_j^*$ of X_j (the point with highest halfspace depth) is the center of gravity of the deepest region (contour) defined as (Rousseeuw and Ruts, 1998)

$$l^*\left(X_j\right) = \max_{\theta_j \in \mathcal{R}^2} \left\{ldepth\left(\theta_j, X_j\right)\right\}, j = 1, ..., k,$$

where this maximum depends on the shape of X_j.

The univariate boxplot is based on ranks since the box goes from the observation with rank $\left\lfloor \frac{n}{4} \right\rfloor$ to that with rank $\left\lceil \frac{3n}{4} \right\rceil$, and the central bar of the box is drawn at the median. A natural generlization of ranks to multivariate data is the notion of halfspace depth (Tukey, 1975). Using this concept in shape analysis, we propose a *bivariate boxplot* (*bagplot*) with the *bag*, that contains 50% of the data points, a *fence* that separates inliers from outliers and a *loop* indicating the landmarks outside the bag but inside the fence (Rousseeuw et al., 1999). The loop plays the same role as the two whiskers in one dimension, so we could call this plot also "*bag-and-bolster plot*" to stress the analogy with "box-and-whisker plot". Like the univariate boxplot, bagplot visualizes several characteristics of the data - location, spread, correlation, skewness and tails. The construction of bag B_j is as follows. Let $\#D_{jl}$ denote the number of data points contained

in D_{jl}. One first determines the value l for which $\#D_{jl} \leq \left\lfloor \frac{n}{2} \right\rfloor < \#D_{j,l-1}$ and then interpolates linearly between D_{jl} and $\#D_{j,l-1}$, relative to $\mathbf{t}_j^*$, to obtain the set B_j (the bug B_j is a convex hull). The fence is obtained by inflating B_j, relative to $\mathbf{t}_j^*$, by a factor 3 found on the base of simulations (Rousseeuw and Ruts, 1997). The loop contains all landmarks between the bag and the fence and its outer boundary is the convex hull of the bag and the non-outliers. The points outside the fence are flagged as outliers.

3. Simulation study

The design of the simulation study is as follows: $n = 1000$ independently randomly generated $k \times 2$ configuration matrices (realizations) $\mathbf{X}_i$ with the rows $\mathbf{x}_{ij}^T$, where $i = 1, ..., 1000$; $j = 1, 2, 3, 4$; $k = 4$, $\mathbf{x}_{ij} = (x_{ij}^{(1)}, x_{ij}^{(2)})^T \sim N_2\left(\mu_{\mathbf{x}_j}, \mathbf{\Sigma}_{\mathbf{x}}\right)$, $\mathbf{\Sigma}_{\mathbf{x}} = \sigma_{\mathbf{x}}^2 \mathbf{I}$, 4×2 matrix $[\mu_{\mathbf{X}}] = [\mu_{\mathbf{x}_1} \vdots \mu_{\mathbf{x}_2} \vdots \mu_{\mathbf{x}_3} \vdots \mu_{\mathbf{x}_4}]^T$, $\mu_{\mathbf{x}_1} = (-1/2, 0)^T$, $\mu_{\mathbf{x}_2} = (1/2, 0)^T$, $\mu_{\mathbf{x}_3} = (0, 1/2)^T$, $\mu_{\mathbf{x}_4} = (0, -1/2)^T$, $\mu_{\mathbf{X}} = Vec([\mu_{\mathbf{X}}])$, 8×8 matrix $\mathbf{\Sigma}_{\mathbf{X}} = \mathbf{I} \otimes \mathbf{\Sigma}_{\mathbf{x}}$ and $\widehat{\sigma}_{\mathbf{x}} = 0.1$. Finally, $Vec(\mathbf{X}_i) \sim N_8(\mu_{\mathbf{X}}, \mathbf{\Sigma}_{\mathbf{X}})$. Using raw data $\mathbf{x}_{ij}$ we calculate Procrustes and Bookstein coordinates with following different choices of baseline end-points: 1) $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$, 2) $\mathbf{x}_{i1}$ and $\mathbf{x}_{i3}$ (Figure 1 and 2).

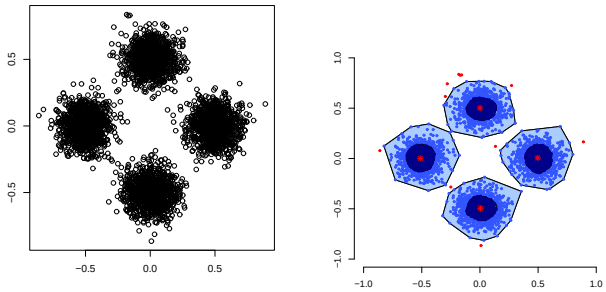


Figure 1: Raw data and bagplots

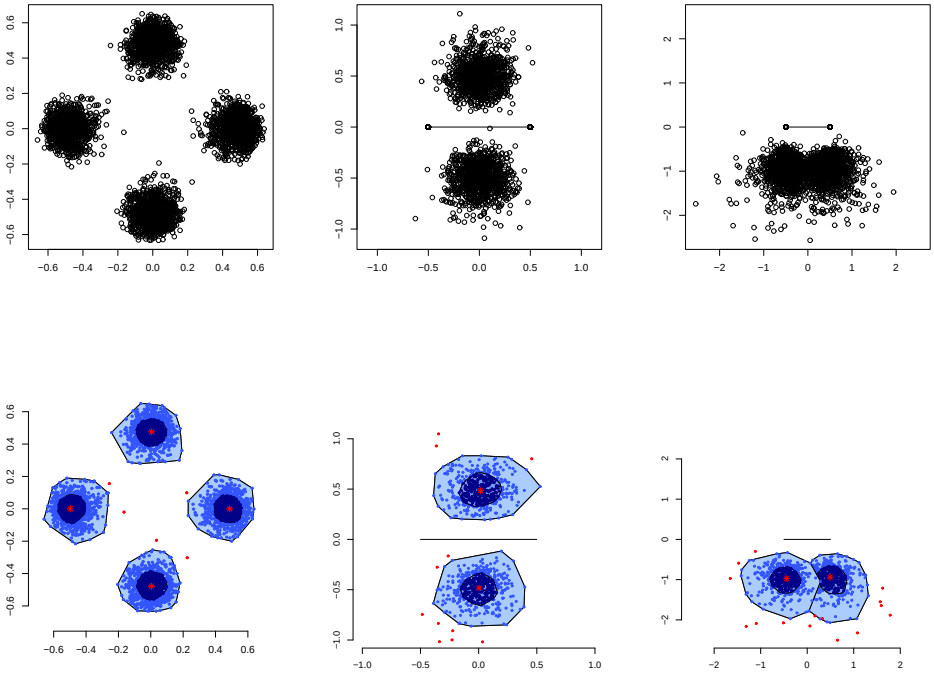


Figure 2: Procrustes and Bookstein coordinates with baseline-choice 1 and 2 (from left to right), data and bagplots

4. Application in biological anthropology

Studied sample consists of 51 female recent human skulls of known age living in the thirties of 20th century. These subjects originated from Pachner collection are located in the Department of Anthropology and Human Genetics, Charles University in Prague (Czech Republic). In total, 20 landmarks (direct marked on the scans of negatives using SigmaScan Pro 5 software) are used in dexter lateral view (Šefčáková et al., 2005). As the baseline for Bookstein coordinates is used the distance of landmarks 11 and 18, also the Procrustes coordinates are used for comparison. The raw data and bagplots of each landmark are plotted to see visual effect of the coordinate-choice (Figure 3).

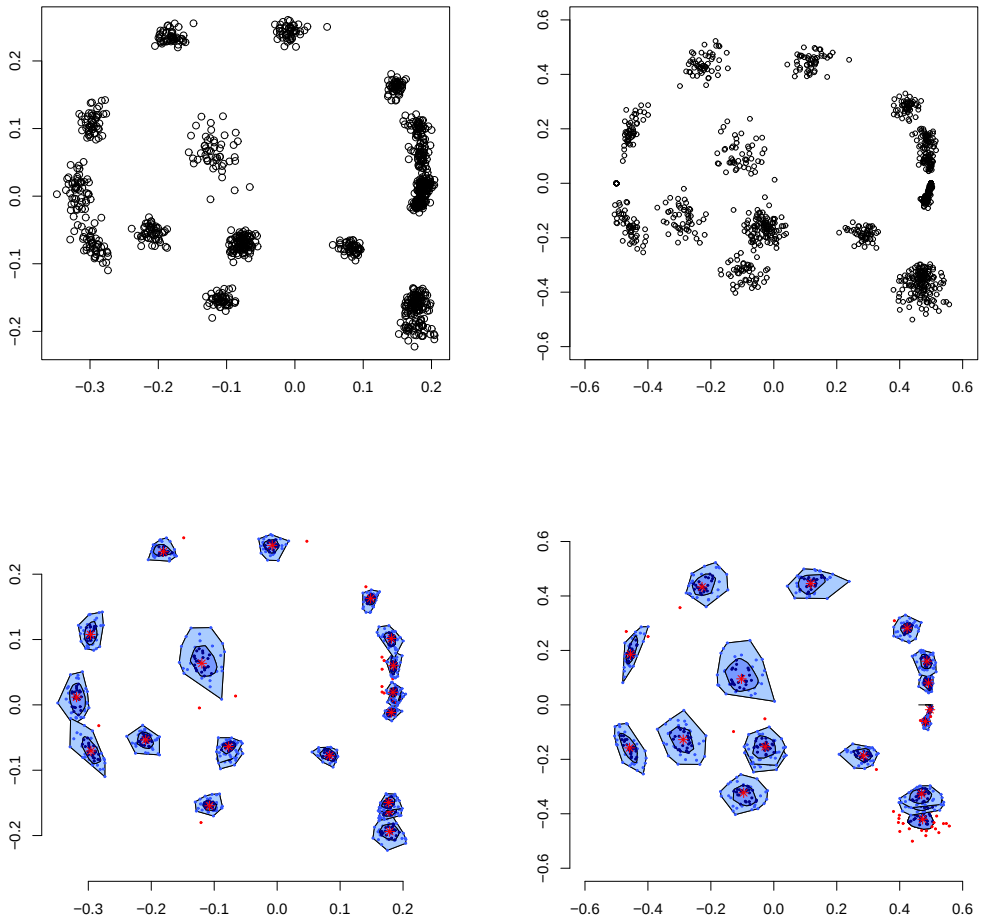


Figure 3: Procrustes and Bookstein coordinates of 51 females from Pachner collection (from left to right), data and bagplots

5. Results and conclusions

From simulations, it is clear that the coordinate-type choice affects the distribution of landmarks (Figure 1). The Procrustes coordinates optimalize position or Procrustes fit between landmarks and minimalize variability of particular landmarks (Figure 2). The Bookstein coordinates are influenced by the choice of baseline. In practice, it is better to use widely separated and less variable landmarks for end-points of baseline. In comparison, taking more close landmarks leads to higher disturbances in distribution of other landmarks (Figure 2). With increasing distance between landmarks and end-points of baseline, the variability of these landmarks is increasing and more outliers can be found (Figure 3). In addition, Procrustes coordinates more reflect real landmarks distribution. In connection with landmarks distribution we propose visualize these distributions by bagplots to see robust Tukey medians and depth contours of landmarks and to detect potential outliers. Other interpretations and applications we let on the reader.

References

1) Dryden, I., Mardia, K. (1999). *Statistical Shape Analysis*. New York: John Wiley and Sons.

2) Goodal, C. (1991). Procrustes Methods in the Statistical Analysis of Shape. *JRSS* **53**, 2: 285 - 339

3) Katina, S. (2004). *Total variation penalty in image warping and selected multivariate methods in shape analysis*. PhD. thesis. Bratislava.

4) Rousseeuw, P.J., Ruts, I. (1997). The bagplot: a Bivariate Box-and-Whisker Plot. (technical report)

5) Rousseeuw, P.J., Ruts, I. (1998). Constructing the bivariate Tukey median. *Statistica Sinica* **8**: 827 - 839

6) Rousseeuw, P.J., Ruts, I., Tukey, J.W. (1999). The bagplot: a Bivariate Boxplot. *The American Statistician* **53**, 4: 382 - 387

7) Šefčáková, A., Katina, S., Brůžek, J., Velemínská, J., Velemínský, P., 2005: Comparison of Gravettian skulls from Předmostí with recent skulls from Pachner collection: roughness penalty approach in shape analysis. *American Journal of Physical Anthropology*. **S40**: 186 - 187

8) Tukey, J.W. (1975). Mathematics and the picturing of data. *Proc. Int. Congress Math., Vancouver* **2**: 523 - 531

9) Ziezold, H., (1994). Mean Figures and Mean Shapes Applied to Biological Figure and Shape Distributions in the Plane. *Biometrical Journal* **36**, 4: 491 - 510

10) Ziezold, H., (2003). *Independence of Landmarks of Shapes*. (preprint)

Acknowledgement: The research was supported by projects No. 436 SLK 17/3/05 (DFG), 1/0272/03 (VEGA) and 206/04/1498 (GAČR). Partially, the research was done while the author was visiting Institute of Mathematics, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany and Department of Mathematics and CS, University of Kassel, Kassel, Germany. In particular, he thanks to Herbert Ziezold for interesting and fruitful discussions and helpful hints, to Alena Šefčáková for partial support with digital measurings, to Jaroslav Brůžek and Petr Velemínský for making digital photos and finally to Jana Velemínská for possibility to use female skulls from Pachner collection in this paper.

PaedDr. RNDr. Stanislav Katina, PhD.
Department of Applied Mathematics and Statistics
Faculty of Mathematics, Physics and CS
Comenius University
Mlynska dolina
842 48 Bratislava
Slovakia
Tel. +421 2 60295135
E-Mail: katina@fmph.uniba.sk
www.defm.fmph.uniba.sk/∼katina/katina.htm/

Algoritmus výberu segmentov plôch
(On an Algorithm of Surface Segment Detection)

Tomáš Kepič, Csaba Török
Department of Mathematics, TU of Košice, Vysokoškolská 4, Slovakia
kepic@acase.sk, csaba.torok@tuke.sk

Abstrakt

Nedávno bol navrhnutý nový prístup APCA k postupnej aproximácii bodov po častiach kubických parabol. V súčasnosti pracujeme na zovšeobecnení metódy pre aproximáciu priestorových dát. Práca popisuje dosiahnuté výsledky pri riešení danej úlohy.

Úvod

Prenosy, uchovávanie a analýzy veľkého množstva dát v súčasnosti vyžadujú nové techniky kompresie a aproximácie. Na riešenie úloh aproximácie a kompresie používame nový prístup založený na základe novej 4-bodovej transformácie – diskrétnej projekčnej transformácie [1].

Práca [2] bola venovaná globálnej funkčnej aproximácii na základe inverznej projekčnej transformácie. Modifikácia použitej techniky nás viedla k jednoduchému modelu, ktorý sa stal využiteľný pri lokálnej 2D aproximácii a umožnil vyvinutie po častiach kubickej aproximácie v automatickom režime - APCA (Autotracking Piecewise Cubic Approximation) [3-4]. APCA nám otvára nové možnosti pre reprezentáciu povrchov vo vizualizačných technikách a kompresii signálu [5]. APCA [3-4] rozdeľuje interval/krivku na podintervaly/segmenty rôznych dĺžok. Na každom podintervale/segmente vykonáva lokálne kubické odhady a dáva techniku pre ziskanie integrálnych kubických aproximantov. Nájdenie pravých (zlomových, hraničných) bodov podintervalov v automatickom režime a iteratívne výpočtové postupy sú dve základné črty navrhovanej metódy využívajúcej špeciálny aproximačný model.

Našim hlavným zámerom je zovšeobecnenie APCA pre aproximáciu 3D dát pomocou plôch. Hlavná myšlienka spočíva v aplikovaní 2D APCA modelu v oboch rovinných smeroch (pozdĺž osi x a pozdĺž osi y) pomocou po častiach vytvorených segmentov

(obdĺžnikov), nad ktorými zostrojíme lokálne aproximačné plochy pomocou efektívneho neúplného kubického regresného modelu.

Pre algoritmizáciu hlavnej úlohy potrebujeme vyriešiť niekoľko podstatných, no podľa našich odhadov, riešiteľných problémov. V práci popisujeme jeden algoritmus riešenia zjednodušenej úlohy.

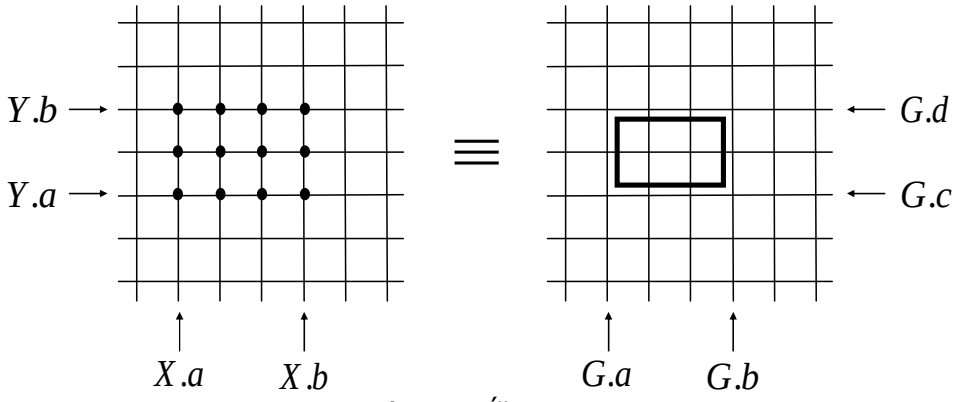
Formulácia zjednodušenej úlohy:
Úlohou je pokryť sieť m x n bodov 2D segmentmi, tak aby platilo:

- segmenty sa navzájom neprekrývajú
- zjednotenie všetkých segmentov pokryje celú pracovnú sieť

2D segment (následnú konštrukciu 3D plochy nad segmentom v danej práci neuvažujeme) sa skonštruuje na základe horizontálnych a vertikálnych intervalov rôznej veľkosti: pravý (horný) koniec intervalu (krivky) sa určuje pomocou kritéria na výber bodov, ktoré môžu byť aproximované kubickými parabolami (pre túto zjednodušenú úlohu pravé a horné body intervalov sa jednoducho generujú).

Definícia základných pojmov

Sieť (m, n) je pravouhlý systém bodov ($m+1$ bodov po osi x a $n+1$ bodov po osi y) s konštantným krokom. Krok h_x (krok pozdĺž osi x) môže byť rôzny od kroku h_y (krok pozdĺž osi y). V práci predpokladáme $h_x = h_y = 1$, teda uvažujeme iba body s celočíselnými súradnicami. BUNV (bez ujmy na všeobecnosti) budeme predpokladať, že ľavý dolný bod siete je [0, 0] (počiatočný bod siete) a pravý horný bod siete je [m, n] (koncový bod).



Obrázok 1: Segment a jeho *vizualizácia* obdĺžnikom

Obdĺžnik (segment) v pravej časti obrázku č.1 zodpovedá množine diskrétnych bodov z ľavého obrázku (znázorňovanie množiny diskrétnych bodov pomocou *„vnútorného"* obdĺžnika je vhodnejšie pre vizualizáciu susedných segmentov vďaka ktorému nedochádza k ich vzájomnému prekrytiu). V práci pod segmentom rozumieme priemet 3D bodov segmentu plochy do roviny x, y (tomuto segmentu zodpovedá v hlavnej úlohe segment aproximačnej plochy).

Poznámka:
Segment znamená:

- *v kontexte plôch buď 3D segment plochy, alebo jeho 2D priemet*
- *v kontexte kriviek buď 2D segment krivky alebo jej priemet.*

Interval $X = \langle a,b \rangle \equiv \langle X.a, X.b \rangle$ je diskrétna množina bodov x:

$$a \le x \le b \quad a \in N,\ x \in N,\ b \in N, a < b\,.$$

Ak X, Y sú intervaly, tak **pseudosegment** $\overline{G}$ definujeme:

$$\overline{G} = \overline{X} \times \overline{Y} \equiv (\overline{X}.a,\ \overline{X}.b \,|\, \overline{Y}.a,\ \overline{Y}.b) \equiv (\overline{G}.a,\ \overline{G}.b \,|\, \overline{G}.c,\ \overline{G}.d)\,.$$

Nasledovne pseudosegment ako časť siete je pravouhlý obdĺžnik, ktorý je kandidátom pre segment (pozri obrázok 1).

Segment G je časť pseudosegmentu: $G \subseteq \overline{G},\ G = X \times Y = (G.a,\ G.b \,|\, G.c,\ G.d)$.

Množinu všetkých segmentov označujeme S.

Poznámka:
V zjednodušenej úlohe pripúšťame dĺžku intervalov X a Y väčšiu ako jedna, teda intervaly môžu obsahovať aj dva body. V hlavnej úlohe však budú musieť intervaly X a Y obsahovať aspoň 4 body z dôvodu dodržania podmienky pre APCA metódu.

Segment G môžeme rozdeliť na **aktívnu časť** G^{akt} a **pasívnu časť** G^{pas}, tak aby platilo: $G = G^{akt} \cup G^{pas},\ G^{akt} \subseteq G,\ G^{pas} \subseteq G,\ G^{akt} \cap G^{pas} = \varnothing$.

Aktívna časť segmentu G^{akt} je tá časť segmentu, ktorá nie je zprava ohraničená iným segmentom, resp. nejakou jeho časťou.

Pasívna časť segmentu G^{pas} je doplnok k aktívnej časti segmentu $G^{pas} = G - G^{akt}$, teda ide o zvyšok, ktorý sa v algoritme ďalej nevyužíva.

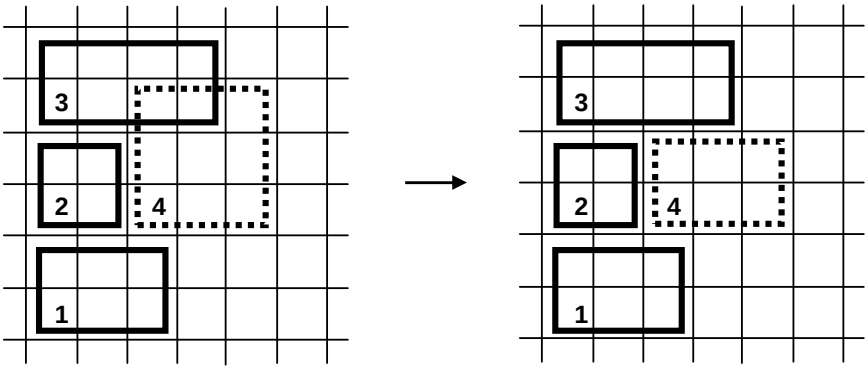
Pravý segment je segment, ktorý obsahuje iba aktívnu časť segmentu (nie je zprava úplne ohraničený).

Pracovná množina segmentov W je množina všetkých pravých segmentov.

Ľavý dolný bod segmentu A (východzí bod segmentu) je bod zo siete, ktorý je počiatočným bodom segmentu $A = [x, y] = [G.a, G.c]$.

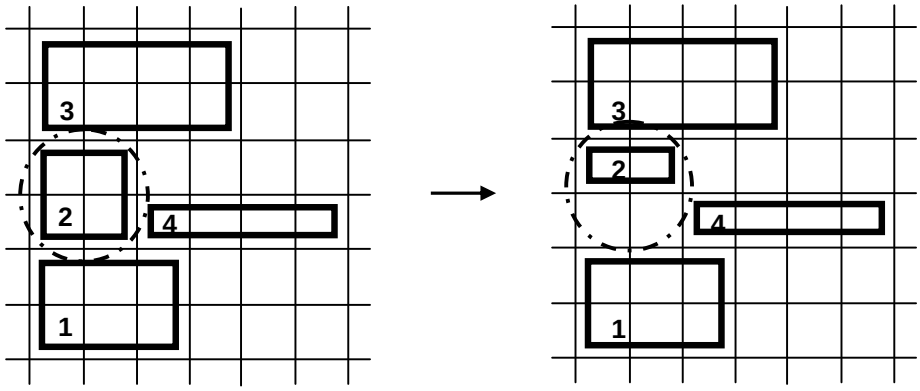
Znovu skonštruovaný segment sa stáva prvkom množiny S a spočiatku je aj pravým segmentom, teda prvkom pracovnej množiny W. V priebehu algoritmu v dôsledku skonštruovania nových segmentov sa pravý segment môže zúžiť, ba dokonca môže úplne vypadnúť z pracovnej množiny W.

Proces zúženia pseudosegmentu je postup, pomocou ktorého získame z pseudosegmentu $\overline{G}$ segment G (pozri obrázok 2).



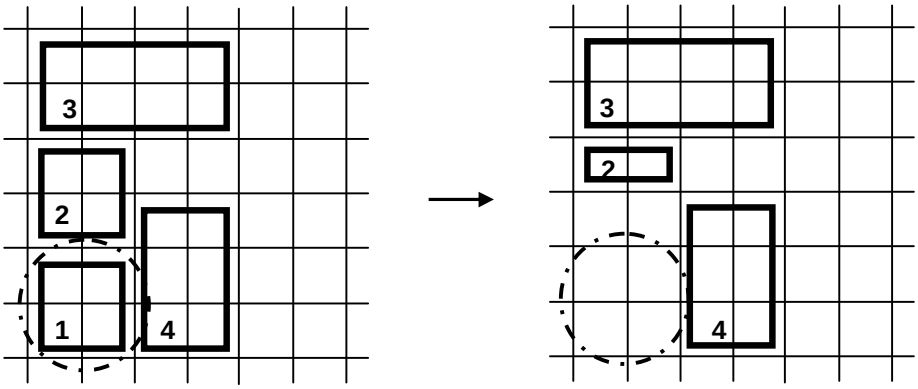
Obrázok 2: Zúženie pseudosegmentu

Proces zúženia segmentu G, je proces rozdelenia segmentu na aktívnu G^{akt} a pasívnu časť G^{pas} (pozri obrázok 3). V pracovnej množine pôvodný segment nahradíme jeho aktívnou časťou.



Obrázok 3: Zúženie segmentu na základe aktívnej časti

Proces zmazania segmentu, je proces odstránenia segmentu z pracovnej množiny segmentov W (pozri obrázok 4).



Obrázok 4: Zmazanie segmentu

Poznámka:
Proces zúženia a zmazania segmentu súvisí iba s pracovnou množinou W, ale nie s množinou segmentov S.

Algoritmus výberu segmentov plôch pre zjednodušenú úlohu

Algoritmus výberu segmentov plôch pozostáva z nasledujúcich krokov:

1) **Inicializácia** – zadanie siete pomocou $(m+1)\times(n+1)$ bodov, $m \in N, n \in N$.
2) **Nájdenie ľavého dolného bodu A_{i+1} pre pseudosegment $\overline{G}_i$**. Nech $A_0 = [0, 0]$ a pre výber bodu A_{i+1}, $i \in N$ používame dve stratégie:

a. stratégia „do hora". Ak platí:

$A_i.y < G_i.d < n \ \wedge \ G_i.a = A_i.x \ \wedge \ A_i.x \geq \max_{j:\, G_j.c > A_i.y} \{G_j.b\}$, tak

$$A_{i+1} = [A_i.x,\, G_i.d] = [G_i.a,\, G_i.d].$$

b. stratégia $\min_{G_j \in W} \{G_j.b\}$ znamená, že kedy spomedzi všetkých segmentov z pracovnej množiny W vyberieme ten, ktorý má minimálne $G_j.b$ (pri rovnosti rozhoduje $\min_{G_j \in \left\{G_k : G_k.b = \min_{G_l \in W}\{G_l.b\}\right\}} \{G_j.c\}$). Na základe tejto stratégie A_{i+1} definujeme nasledovne:

$$A_{i+1} = \left[\min_{G_j \in W} \{G_j.b\}, \min_{G_j \in \left\{G_k : G_k.b = \min_{G_l \in W}\{G_l.b\}\right\}} \{G_j.c\} \right].$$

3) **Generovanie pseudosegmentu** $\overline{G}_i$. 2D segment skonštruujeme na základe horizontálnych a vertikálnych 1D intervalov. Pomocou 2D APCA môžeme detekovať pravý (horný) koniec segmentu. V zjednodušenej úlohe pravé konce intervalu generujeme pomocou pseudonáhodných čísel.

4) **Pri aplikovaní procesu zúženia pseudosegmentu** $\overline{G}_i$, výsledkom ktorého dostávame segment G_i, ide o zúženie časti pseudosegmentu o tú časť, ktorá sa prekrýva s inými segmentmi:

$$G_i = \left(\overline{G}_i.a,\, \overline{G}_i.b \,|\, \overline{G}_i.c,\, \min\left\{ \overline{G}_i.d,\, \min_{j:\overline{G}_i.c < G_j.c < \overline{G}_i.d,\, i \neq j} \{G_j.c\} \right\} \right).$$

5) **Testovanie podmienky ukončenia algoritmu výberu**. Podmienku ukončenia definujeme: $G_i.b = m \ \wedge \ G_i.d = n$. Algoritmus sa ukončí pri splnení tejto podmienky.

6) **Aplikovanie procesov zúženia a zmazania na základe segmentu** G_i.

a. V procese zmazania po pridaní segmentu G_i do pracovnej množiny W sa z nej zmažú všetky segmenty G_j, ktoré obsahujú iba pasívnu časť. Podmienka zmazania segmentov G_j je: $G_j.d \leq G_i.d \ \wedge \ G_j.b \leq G_i.a$

b. Podmienka zúženia segmentov:

$G_j : G_j.b = G_i.a \ \wedge \ G_j.c < G_i.d \ \Rightarrow \ G_j = \left(G_j.a,\, G_j.b \,|\, G_i.d,\, G_j.d\right)$ sa

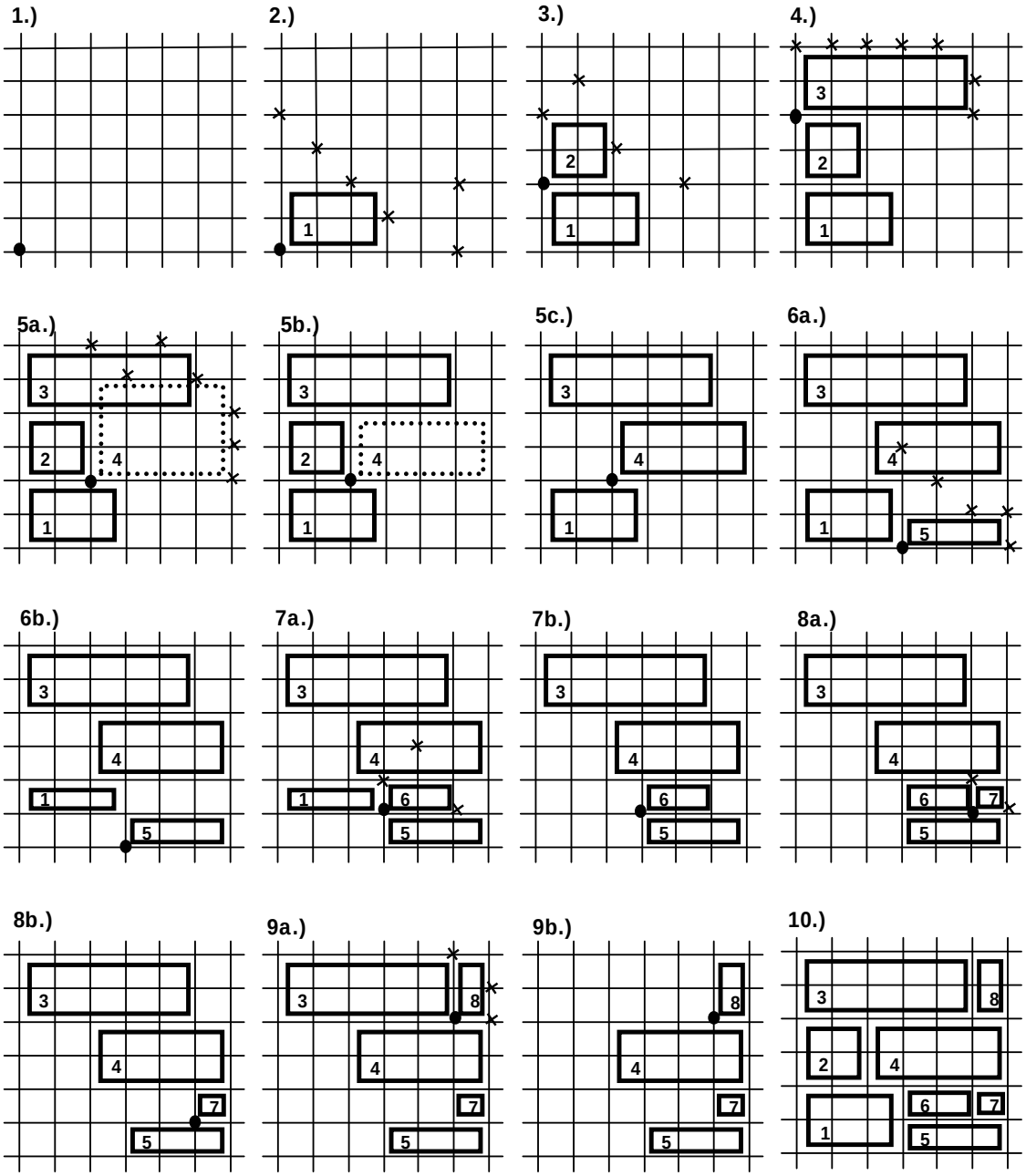
používa v procese zúženia segmentov z pracovnej skupiny W po pridaní nového segmentu G_j do nej.

7) $i = i + 1$, pokračuj bodom 2)

Algoritmus výberu segmentov plôch ilustrujeme príkladom (pozri obrázok 5), ktorému zodpovedá iteračná tabuľka (pozri tabuľku 1):

Obr.	Postupnosť generovaných bodov	Východzí bod	Pracovná množina W	Množina segmentov S
1		A = [0, 0]	W = Ø	S = Ø
2	{ 5, 4, 3, 3, 5, 3 }	A = [0, 0]	W = { (0, 3|0,2) }	S = { (0, 3|0,2) }
3	{ 4, 4, 2, 5 }	A = [0, 2]	W = { (0, 3|0,2), (0, 2|2, 4) }	S = { (0, 3|0,2), (0, 2|2, 4) }
4	{ 5, 6, 5, 6, 6, 6, 6 }	A = [0, 4]	W = { (0, 3|0,2), (0, 2|2, 4), (0, 5|4, 6) }	S = { (0, 3|0,2), (0, 2|2, 4), (0, 5|4, 6) }
5a	{ 6, 6, 6, 5, 6, 6, 5 }	A = [2, 2]		S = { (0, 3|0,2), (0, 2|2, 4), (0, 5|4, 6) }
5b		A = [2, 2]	W = { (0, 3|0,2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4) }	S = { (0, 3|0,2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4) }
5c		A = [2, 2]	W = { (0, 3|0,2), (0, 5|4, 6), (2, 6|2, 4) }	S = { (0, 3|0,2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4) }
6a	{ 6, 3, 6, 2, 1 }	A = [3, 0]	W = { (0, 3|0, 2), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1) }	S = { (0, 3|0, 2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1) }
6b		A = [3, 0]	W = { (0, 3|1, 2), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1) }	S = { (0, 3|0, 2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1) }
7a	{ 5, 2, 3 }	A = [3, 1]	W = { (0, 3|1, 2), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (3, 5|1, 2) }	S = { (0, 3|0, 2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (3, 5|1, 2) }
7b		A = [3, 1]	W = { (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (3, 5|1, 2) }	S = { (0, 3|0, 2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (3, 5|1, 2) }
8a	{ 6, 2 }	A = [5, 1]	W = { (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (3, 5|1, 2), (5, 6|1, 2) }	S = { (0, 3|0, 2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (3, 5|1, 2), (5, 6|1, 2) }
8b		A = [5, 1]	W = { (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (5, 6|1, 2) }	S = { (0, 3|0, 2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (3, 5|1, 2), (5, 6|1, 2) }
9a	{ 6, 6, 6 }	A = [5, 4]	W = { (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (5, 6|1, 2), (5, 6|4, 6) }	S = { (0, 3|0, 2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (3, 5|1, 2), (5, 6|1, 2), (5, 6|4, 6) }
9b		A = [5, 4]	W = { (2, 6|2, 4), (3, 6|0, 1), (5, 6|1, 2), (5, 6|4, 6) }	S = { (0, 3|0, 2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (3, 5|1, 2), (5, 6|1, 2), (5, 6|4, 6) }
10				S = { (0, 3|0,2), (0, 2|2, 4), (0, 5|4, 6), (2, 6|2, 4), (3, 6|0, 1), (3, 5|1, 2), (5, 6|1, 2), (5, 6|4, 6) }

Tabuľka 1: Iteračná tabuľka k ilustrovanému príkladu výberu segmentov plôch



Obrázok 5: Ilustrovaný príklad výberu segmentov plôch

Záver

V práci sme popísali algoritmus formulovanej úlohy, ktorá je časťou hlavnej úlohy aproximovať 3D body kubickými segmentmi na základe APCA kritéria. V budúcnosti plánujeme:

- zjednodušiť daný algoritmus výberu segmentov o stratégiu „do hora" pri hľadaní nového bodu A_{i+1} pre pseudosegment $\overline{G}_i$.
- zovšeobecniť zjednodušenú úlohu pre prípad, keď intervaly X, Y obsahujú najmenej 4 body
- nahradiť generátor pseudonáhodných čísel pomocou 2D APCA kritéria ako v smere osi x, tak aj v smere osi y.
- nad 2D segmentom aproximovať 3D body pomocou neúplného bikubického modelu IBM (Incomplete Bicubic Model).

Literatúra

[1] Dikoussar N.D.: Function parameterization by using 4-point transforms, Comput. Phys. Commun. 99 (1997), 235-254

[2] Török Cs.: Dikoussar N.D.: Approximation with DPT, Computers and Mathematics with Applications, 38 (1999) p.211-220

[3] Dikoussar N.D., Török Cs.: Piecewise Cubic Approximation in Autotracking Mode, 2004, Communication JINR Dubna, to appear

[4] Török Cs., Dikoussar N.D.: MS .NET components for piecewise cubic approximation, Communication JINR, Dubna, 2004, to appear

[5] Kepič T., Török Cs., Dikoussar N.D.: Wavelet compression, 13. International Workshop on computational statistics, Bratislava, 2004, pp.49-52

Porovnanie Slangerových odhadov variančných komponentov so štandardnými odhadmi pomocou simulácií

Daniel Klein

ABSTRAKT: The main goal of this paper is to compare the new method of estimating variance components, introduced by Slanger, with widely used methods, such as restricted maximum likelihood and Henderson 3 method.

1 ÚVOD

Slanger vo svojom článku [2] prezentoval novú metódu odhadu variančných komponentov v modeloch so zmiešanými efektami. Ako autor uviedol, táto metóda má isté výhody oproti bežne používaným metódam: nie sú potrebné žiadne apriórne hodnoty neznámych parametrov, nie je to metóda iteračná a je použiteľná v akomkoľvek lineárnom modeli so zmiešanými efektami. Slanger & Carlson [3] urobili simulačnú štúdiu, kde porovnávali túto novú metódu s odhadmi metódou reziduálnej maximálnej vierohodnosti. Výsledok hovoril v prospech Slangerovej metódy.

Avšak niektoré aspekty tejto metódy sa zdajú byť podozrivé:

- Lehmann a Scheffé (1950) dokázali, že odhad *T* funkcie $h(\Theta)$ je rovnomerne najlepší odhad v triede nestranných odhadov funkcie $h(\Theta)$ práve vtedy, ak odhad *T* je nekorelovaný s každou štatistikou *Z*, ktorej stredná hodnota je nulová. Slanger vo svojej metóde obmedzil množinu všetkých takýchto štatistík len na štatistiky v tvare kvadratickej formy s nulovou strednou hodnotou.
- K získaniu odhadov je nutné riešiť systém lineárnych rovníc. Ak existuje riešenie tohto systému, kritérium nekorelovanosti bolo splnené a odhad nájdený. V opačnom prípade systém sa vyrieši metódou najmenších štvorcov, čím sa získa odhad variančného komponentu z geometrického hľadiska najbližšie k splneniu podmienky nekorelovanosti. To však nezaručuje žiadne štatistické vlastnosti.

To nás viedlo k simulačnej štúdii, pričom odhady získané touto metódou sme porovnávali s odhadmi metódou maximálnej vierohodnosti (ML), reziduálnej maximálnej vierohodnosti (REML) a odhadmi získanými Hendersonovou metódou 3 (H3).

2 PRINCÍP METÓDY

Slanger odvodil túto metódu v jednoduchom modeli s dvoma variančnými komponentmi:

$$Y = X\beta + Z\alpha + \varepsilon,$$

kde β sú fixné a α náhodné efekty. Ak $\mathrm{var}(\varepsilon) = \sigma_\varepsilon^2 I$ a $\mathrm{var}(\alpha) = \sigma_\alpha^2 A$, potom $\mathrm{E}(Y) = X\beta$ a $\mathrm{var}(Y) = V = \sigma_\varepsilon^2 I + \sigma_\alpha^2 ZAZ'$. Pre zjednodušenie označme $J = ZAZ'$. Potom $V = \sigma_\varepsilon^2 I + \sigma_\alpha^2 J$. Variančné komponenty sú odhadované kvadratickými formami $\hat{\sigma}_\varepsilon^2 = Y'M_X Q_\varepsilon M_X Y$ a $\hat{\sigma}_\alpha^2 = Y'M_X Q_\alpha M_X Y$, kde $M_X = I - P_X = I - X(X'X)^- X'$. Pretože

$$\mathrm{E}(Y'M_X QM_X Y) = \mathrm{Tr}(QM_X VM_X) = \sigma_\alpha^2 (\mathrm{vec}(M_X JM_X))' \mathrm{vec}(Q) + \sigma_\varepsilon^2 (\mathrm{vec}(M_X))' \mathrm{vec}(Q),$$

aby odhady boli nestranné, musia *Q* matice spĺňať

$$(\mathrm{vec}(M_X JM_X))' \mathrm{vec}(Q_\alpha) = 1 \quad \text{a} \quad (\mathrm{vec}(M_X))' \mathrm{vec}(Q_\alpha) = 0,$$
$$(\mathrm{vec}(M_X JM_X))' \mathrm{vec}(Q_\varepsilon) = 0 \quad \text{a} \quad (\mathrm{vec}(M_X))' \mathrm{vec}(Q_\varepsilon) = 1.$$

Aby odhady podľa Lehmann-Scheffé lemmy boli rovnomerne najlepšími odhadmi, musia byť nekorelované s každým odhadom nuly. Nech $T = Y'M_X SM_X Y$ je štatistika, ktorá je odhadom nuly, t. z. $\mathrm{E}(T) = \mathrm{E}(Y'M_X SM_X Y) = 0$. Z nulovosti strednej hodnoty potom vyplýva, že matica *S* musí spĺňať

$$\begin{pmatrix} (\mathrm{vec}(M_X))' \\ (\mathrm{vec}(M_X JM_X))' \end{pmatrix} \mathrm{vec}(S) \overset{\mathrm{df}}{=} B\,\mathrm{vec}(S) = 0\,.$$

To znamená, že $\mathrm{vec}(S)$ musí ležať v nulovom priestore matice *B*. V ďalšom stačí pracovať s bázou tohto priestoru, pretože akýkoľvek prvok je možné získať ako lineárnu kombináciu prvkov z bázy. Ak teda $\mathrm{vec}(C_1),\ldots,\mathrm{vec}(C_m)$ tvoria bázu nulového priestoru matice *B*, potom $S = \sum_{i=1}^{m} k_i C_i$.

Kovariancia medzi dvoma kvadratickými formami $Y'M_X QM_X Y$ a T , ak Y je normálne rozdelený, je

$$\begin{aligned}\mathrm{cov}(Y'M_X QM_X Y, Y'M_X SM_X Y) &= 2\sigma_\alpha^4 (\mathrm{vec}(M_X JM_X SM_X JM_X))'\,\mathrm{vec}(Q) + \\ &+ 2\sigma_\alpha^2\sigma_\varepsilon^2 (\mathrm{vec}(M_X JM_X SM_X + M_X SM_X JM_X))'\,\mathrm{vec}(Q) + \\ &+ 2\sigma_\varepsilon^4 (\mathrm{vec}(M_X SM_X))'\,\mathrm{vec}(Q)\,.\end{aligned}$$

Aby bola kovariancia nulová, musí pre $\forall i$ platiť

$$\begin{gathered}(\mathrm{vec}(M_X JM_X C_i M_X JM_X))'\,\mathrm{vec}(Q) = 0\,, \\ (\mathrm{vec}(M_X JM_X C_i M_X + M_X C_i M_X JM_X))'\,\mathrm{vec}(Q) = 0\,, \\ (\mathrm{vec}(M_X C_i M_X))'\,\mathrm{vec}(Q) = 0\,.\end{gathered}$$

Ak si označíme $C_v = (\mathrm{vec}(C_1),\ldots,\mathrm{vec}(C_m))$, potom predchádzajúce podmienky môžeme prepísať

$$\begin{pmatrix} C_v'(M_X JM_X \otimes M_X JM_X) \\ C_v'(M_X JM_X \otimes M_X + M_X \otimes M_X JM_X) \\ C_v'(M_X \otimes M_X) \end{pmatrix} \mathrm{vec}(Q) \overset{\mathrm{df}}{=} R_0\,\mathrm{vec}(Q) = 0\,.$$

Označme si ďalej e_i vektor obsahujúci 1 na *i*-tom mieste, inde 0. Záporné *i* znamená indexovanie odzadu, t. z. e_{-1} zanmená vektor, ktorý má 1 na poslednom mieste.

Odhad variančného komponentu potom získame riešením sústavy lineárnych rovníc

$$\begin{pmatrix} B \\ R_0 \end{pmatrix} \mathrm{vec}(Q) \overset{\mathrm{df}}{=} R\,\mathrm{vec}(Q) = h,$$

kde $h = e_2$ pre Q_α a $h = e_1$ pre Q_ε. Riešenie metódou najmenších štvorcov je $\mathrm{vec}(Q) = (R'R)^{-}R'h$. Ak $R\,\mathrm{vec}(Q) - h = 0$, Lehmann-Scheffého kritérium bolo podľa autora splnené a rovnomerne najlepší nestranný odhad bol nájdený. V opačnom prípade, ak sa použije Moore-Penroseova pseudoinverzia, získa sa odhad, ktorý je z geometrického hľadiska najbližšie k splneniu Lehmann-Scheffého kritéria. Nie je problém zovšeobecniť tento postup na akýkoľvek iný lineárny model so zmiešanými efektami.

V niektorých situáciách však takýmto postupom môže dôjsť k porušeniu nestrannosti. Preto Slanger navrhol použiť metódu Lagrangeových multiplikátorov na zabezpečenie nestrannosti. Označenia na rozlíšenie týchto dvoch prístupov budú $\mathrm{LSLS_b}$ (prezentovaný vyššie) a $\mathrm{LSLS_u}$ (prezentovaný v ďalšom).

Pre model s dvoma variančnými komponentmi bude na získanie $LSLS_u$ odhadu mať systém lineárnych rovníc tvar

$$\begin{pmatrix} R_0' R_0 & B' \\ B & 0 \end{pmatrix} \begin{pmatrix} \mathrm{vec}(Q) \\ \lambda \end{pmatrix} = h,$$

kde $h = e_{-1}$ pre Q_α, $h = e_{-2}$ pre Q_ε a λ sú Lagrangeove multiplikátory.

3 SIMULÁCIE

Pre simulácie sme uvažovali 2 modely:

a) model so zmiešanými efektami,
b) model s náhodnými efektami.

a) Použili sme lineárny model so zmiešanými efektami, dvojité triedenie (viď [1])

$$y_{ijk} = \mu + \alpha_i + \eta_j + \gamma_{ij} + \varepsilon_{ijk},$$

$i = 1,2$, $j = 1,2,3$, $k = 1,\ldots,n_{ij}$, kde $\mathrm{E}(\eta_j) = \mathrm{E}(\gamma_{ij}) = \mathrm{E}(\varepsilon_{ijk}) = 0$, $\mathrm{var}(\varepsilon_{ijk}) = \sigma_\varepsilon^2 = 64$, $\mathrm{var}(\eta_j) = \sigma_\eta^2 = 784$, $\mathrm{var}(\gamma_{ij}) = \sigma_\gamma^2 = 25$. Všetky náhodné efekty boli uvažované nezávislé a normálne rozdelené. Fixné efekty boli uvažované $\mu + \alpha_1 = 200$ a $\mu + \alpha_2 = 150$. Počet pozorovaní v jednotlivých skupinách ukazujú tabuľky 1. a 2. V prvom modeli všetky bunky obsahujú pozorované dáta, v druhom sú aj nejaké chýbajúce dáta (budeme používať označenia model 1 a model 2 pre jednotlivé modely).

TABUĽKA 1

n_{ij}				$n_{i.}$
	3	5	5	13
	6	3	3	12
$n_{.j}$	9	8	8	n=25

TABUĽKA 2

n_{ij}				$n_{i.}$
	3	8	7	18
	-	4	3	7
$n_{.j}$	3	12	10	n=25

b) Pre model s náhodnými efektami sme použili model z predchádzajúceho, kde α sme uvažovali ako náhodný efekt. Počet pozorovaní v jednotlivých skupinách ukazuje tabuľka 3. Všetky náhodné efekty boli uvažované nezávislé a normálne rozdelené, pričom $\mathrm{E}(\alpha_i) = \mathrm{E}(\eta_j) = \mathrm{E}(\gamma_{ij}) = \mathrm{E}(\varepsilon_{ijk}) = 0$, $\mathrm{var}(\varepsilon_{ijk}) = \sigma_\varepsilon^2 = 64$, $\mathrm{var}(\alpha_i) = \sigma_\alpha^2 = 160$, $\mathrm{var}(\eta_j) = \sigma_\eta^2 = 784$, $\mathrm{var}(\gamma_{ij}) = \sigma_\gamma^2 = 25$. Spoločná stredná hodnota bola uvažovaná $\mu = 150$.

TABUĽKA 3

n_{ij}				$n_{i.}$
	3	5	5	13
	6	3	3	12
$n_{.j}$	9	8	8	n=25

Vo všetkých prípadoch bolo generovaných 5000 vektorov pozorovaní a variančné komponenty boli vypočítané uvažovanými metódami. K simuláciám boli použité balíky R.2.0.0 a SPSS 10.1.

4 ZÁVER

Napriek tomu, že simulačná štúdia, ktorú urobili Slanger&Carlson, vyznieva v prospech Slangerovej metódy oproti metóde reziduálnej maximálnej vierohodnosti a podľa autorov je to nová metóda, ktorá sľubuje vyššiu presnosť ako ktorákoľvek iná metóda [3], naše simulácie tento výsledok nepotvrdili.

Jedine odhad reziduálnej chyby všetkými uvažovanými metódami bol vo všetkých modeloch porovnateľný. Odhady boli pomerne presné, okrem odhadu metódou $LSLS_u$ v modeli s náhodnými efektami. V tomto prípade mal odhad neporovnateľne vyšší rozptyl oproti ostatným metódam.

Pre odhad ostatných variančných komponentov bola LSLS metóda veľmi nespoľahlivá. Stredné hodnoty tejto metódy neboli tak presné a blízke skutočným hodnotám, ako to bolo u metód REML a H3. Slangerova metóda však mala zvyčajne menej záporných odhadov ako zvyšné metódy.

Pretože sme poznali skutočné hodnoty parametrov, mohli sme vypočítať očakávané stredné hodnoty odhadov variančných komponentov pre metódu $LSLS_b$ vo všetkých uvažovaných situáciách (pozri tabuľku 4). Je vidieť, že len očakávaná stredná hodnota odhadov reziduálnych variančných komponentov je blízka skutočným hodnotám. V ostatných prípadoch sú očakávané stredné hodnoty odhadov podhodnotené, príp. nadhodnotené.

TABUĽKA 4

	$E(\hat{\sigma}_\alpha^2)$ $\sigma_\alpha^2 = 160$	$E(\hat{\sigma}_\eta^2)$ $\sigma_\eta^2 = 784$	$E(\hat{\sigma}_\gamma^2)$ $\sigma_\gamma^2 = 25$	$E(\hat{\sigma}_\varepsilon^2)$ $\sigma_\varepsilon^2 = 64$
Model 1	-	445,7678	21,224	63,908
Model 2	-	149,7235	144,405	64,4362
Model 3	86,7077	596,6294	45,71	64,4868

V dôsledku týchto výsledkov sa táto nová metóda nezdá byť tak dobrá a presná, ako to bolo prezentované autorom.

REFERENCIE

[1] Christensen, R.: *Plane answers to complex question. The theory of linear models.* Springer-Verlag New York Inc., 1987, 243.

[2] Slanger, W. D.: *Least Squares Lehmann-Scheffé estimation of variances and covariances with any mixed linear model.* Journal of Animal Science, 74(11), 2577-2585, 1996.

[3] Slanger, W. D., Carlson, J.: *A comparison via simulation of least squares Lehmann-Scheffé estimators of two variances and heritability with those of restricted maximum likelihood.* Journal of Animal Science, 81(8), 1950-1958, 2003.

Mgr. Daniel Klein
Ústav matematických vied, UPJŠ, Košice
danoklein@hotmail.com

Opisná štatistika vybraných absolútnych ukazovateľov stavebných podnikov v SR

Samuel Korónyi

Abstract: The paper provides basic descriptive statistics of some absolute economic production indicators of Slovak private construction companies from the year 2001 (output: value added, inputs: sum of assets and average number of employee) including graphical presentation in boxplots.

Úvod

Cieľom príspevku je uvedenie základných štatistických parametrov vybraných absolútnych produkčných ukazovateľov súkromných slovenských stavebných podnikov právnej formy a. s. a s. r. o. Údaje sú z výkazu ŠÚ SR Prod 3-04 za rok 2001. Výstupný ukazovateľ je pridaná hodnota (v mil. Sk), vstupný: pracovná sila – priemerný evidenčný počet zamestnancov vo fyzických osobách a kapitál – celkový kapitál (v mil. Sk). Súbor tvoria všetky podniky na Slovensku s klasifikáciou OKEČ 45, ktoré poslali vyplnený výkaz. Vzhľadom na prakticky stopercentnú návratnosť je možné považovať súbor za základný.

Výsledky

Prvým ukazovateľom je priemerný evidenčný počet zamestnancov vo fyzických osobách. Tabuľka 1 uvádza štatistické parametre ukazovateľa priemerného evidenčného počtu zamestnancov vo fyzických osobách pre stavebné podniky právnej formy a. s. a s. r. o.

Tabuľka 1

Porovnanie štatistických parametrov ukazovateľa priemerného evidenčného počtu zamestnancov vo fyzických osobách podľa právnej formy stavebných podnikov

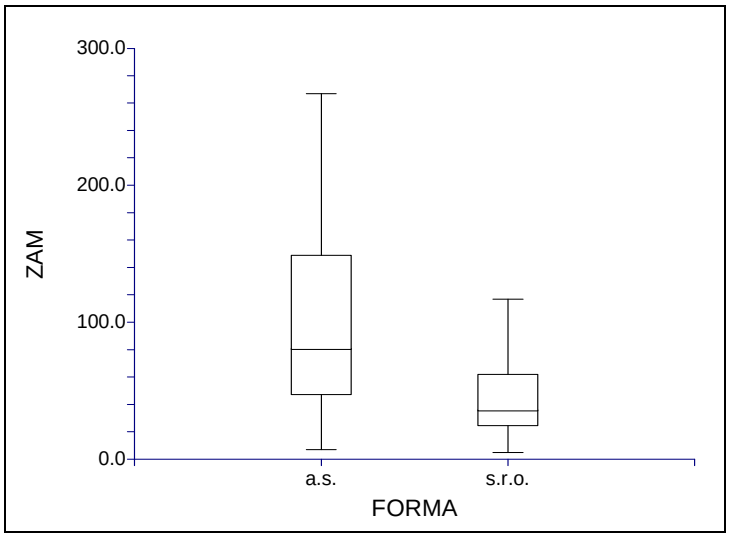
Parameter	a. s.	s. r. o.
N	105	361
μ	132	43
$x_{0,5}$	87	35
σ	140,679	25,683
V	1,062	0,601
Minimum	7	5
Maximum	780	149
$\sqrt{\beta_1}$	2,769	1,257
β_2	8,921	1,284

N = rozsah súboru, μ = aritmetický priemer, $x_{0,5}$ = medián, σ = smerodajná odchýlka, V = variačný koeficient, $\sqrt{\beta_1}$ = parameter šikmosti, β_2 = parameter špicatosti

Interpretácia výsledku:

Stredná hodnota priemerného evidenčného počtu zamestnancov vo fyzických osobách je vyššia v prípade podnikov právnej formy a. s. (aritmetický priemer = 132, medián = 87 zamestnancov) voči forme s. r. o. (aritmetický priemer = 43, medián = 35 zamestnancov). Smerodajná odchýlka priemerného evidenčného počtu zamestnancov vo fyzických osobách je vyššia v prípade podnikov právnej formy a. s.(141) voči forme s. r. o. (26). Variačný koeficient priemerného evidenčného počtu zamestnancov vo fyzických osobách je vyšší v prípade podnikov právnej formy a. s. (106 %) voči forme s. r. o. (60 %). Parameter šikmosti a parameter špicatosti priemerného evidenčného počtu zamestnancov vo fyzických

osobách je vyšší v prípade podnikov právnej formy a. s. (2,8 verzus 1,3) resp. (8,9 verzus 1.3) voči podnikom formy s. r. o. Z tabuľky (hlavne z variačných koeficientov a parametrov šikmosti a špicatosti) je zrejmé, že rozdelenia priemerného evidenčného počtu zamestnancov vo fyzických osobách nie sú normálne. Na grafe 1 je vo forme škatuľového diagramu zobrazené porovnanie priemerného evidenčného počtu zamestnancov vo fyzických osobách podľa právnej formy stavebných podnikov.



Graf 1 Porovnanie priemerného evidenčného počtu zamestnancov (ZAM) vo fyzických osobách podľa právnej formy podniku

Ďalším skúmaným ukazovateľom je celkový úhrn aktív. Tabuľka 2 uvádza štatistické parametre celkového úhrnu aktív pre stavebné podniky právnej formy a. s. a s. r. o.

Tabuľka 2

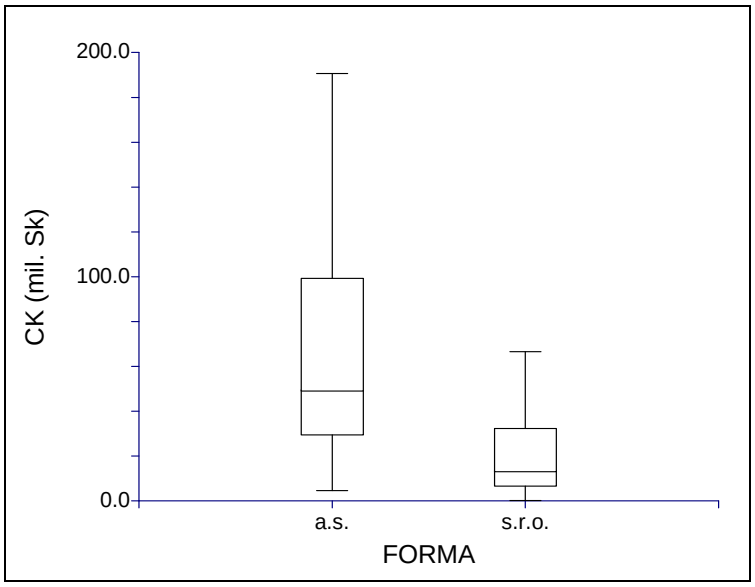
Porovnanie štatistických parametrov ukazovateľa celkového úhrnu aktív (v mil. Sk) podľa právnej formy stavebných podnikov

Parameter	**a. s.**	**s. r. o.**
N	105	361
μ	106,160	21,363
$x_{0,5}$	62,815	12,604
Σ	122,957	24,725
V	1,158	1,157
Minimum	4,557	0,213
Maximum	632,787	196,010
$\sqrt{\beta_1}$	2,384	2,744
β_2	6,064	10,375

Interpretácia výsledku:

Stredná hodnota celkového úhrnu aktív je vyššia v prípade podnikov právnej formy a. s. (aritmetický priemer = 106 mil. Sk, medián = 63 mil. Sk) voči forme s. r. o. (aritmetický priemer = 21 mil. Sk, medián = 13 mil. Sk). Smerodajná odchýlka celkového úhrnu aktív je vyššia v prípade podnikov právnej formy a. s. (123 mil. Sk) voči forme s. r. o. (25 mil. Sk). Variačný koeficient celkového úhrnu aktív je rovnaký pre podniky obidvoch právnych foriem

(116 %). Parameter šikmosti a parameter špicatosti celkového úhrnu aktív je vyšší v prípade podnikov právnej formy s. r. o. (2,7 verzus 2,4) resp. (10,4 verzus 6,1) voči forme a. s. Z tabuľky (hlavne z variačných koeficientov a parametrov šikmosti a špicatosti) je zrejmé, že rozdelenia celkového úhrnu aktív nie sú normálne. Na grafe 3.2 je vo forme škatuľového diagramu zobrazené porovnanie celkového úhrnu aktív podľa právnej formy stavebných podnikov.



Graf 2 Porovnanie celkového úhrnu aktív (CK) podnikov podľa právnej formy

Výstupným ukazovateľom produkčných funkcií je pridaná hodnota. Tabuľka 3 uvádza štatistické parametre ukazovateľa pridanej hodnoty (v mil. Sk) pre stavebné podniky právnej formy a. s. a s. r. o.

Tabuľka 3

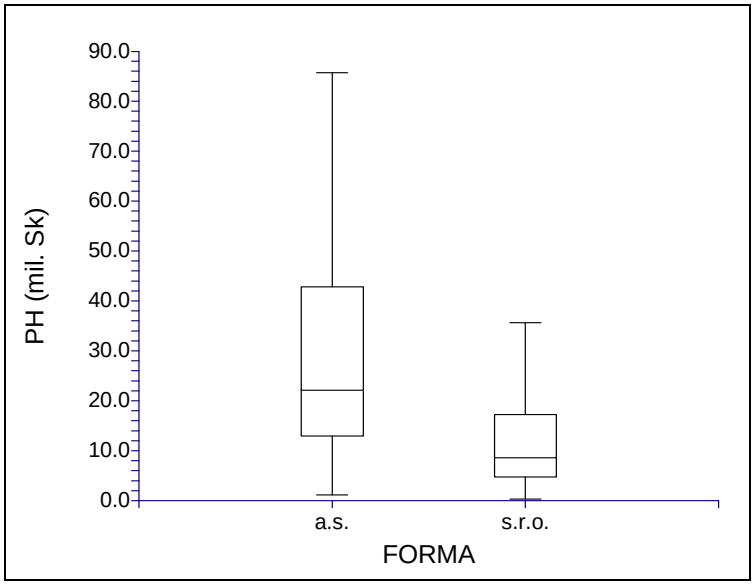
Porovnanie štatistických parametrov ukazovateľa pridanej hodnoty podľa právnej formy stavebných podnikov

Parameter	**a. s.**	**s. r. o.**
N	105	361
μ	38,450	12,113
$x_{0,5}$	23,192	8,375
σ	45,451	11,431
V	1,182	0,944
Minimum	1,145	0,301
Maximum	255,150	81,200
$\sqrt{\beta_1}$	2,892	2,495
β_2	9,141	9,021

Interpretácia výsledku:

Stredná hodnota pridanej hodnoty je vyššia v prípade podnikov právnej formy a. s. (aritmetický priemer = 38 mil. Sk, medián = 23 mil. Sk) voči forme s. r. o. (aritmetický priemer = 12 mil. Sk, medián = 8 mil. Sk). Smerodajná odchýlka pridanej hodnoty je vyššia v prípade podnikov právnej formy a. s. (45 mil. Sk) voči forme s. r. o. (11 mil. Sk). Variačný koeficient pridanej hodnoty je vyšší v prípade podnikov právnej formy a. s. (118 %) voči

forme s. r. o. (94 %). Parameter šikmosti a parameter špicatosti pridanej hodnoty je vyšší v prípade podnikov formy a. s. (2,9 verzus 2,5) resp. (9,1 verzus 9,0) voči podnikom formy s. r. o. Z tabuľky (hlavne z variačných koeficientov a parametrov šikmosti a špicatosti) je zrejmé, že rozdelenia pridanej hodnoty nie sú normálne. Na grafe 3 je vo forme škatuľového diagramu zobrazené porovnanie pridanej hodnoty podľa právnej formy stavebných podnikov.



Graf 3 Porovnanie pridanej hodnoty (PH) podnikov podľa právnej formy

Záver

Stavebné podniky právnej formy a. s. so súkromným tuzemským vlastníctvom majú v porovnaní s podnikmi formy s. r. o. vyššie skúmané absolútne ukazovatele v parametroch strednej hodnoty: priemerný evidenčný počet zamestnancov vo fyzických osobách, celkový úhrn aktív, pridanú hodnotu. Tiež je zrejmá silná nenormalita rozdelenia ukazovateľov.

Použitá literatúra

Hintze, J.L.: NCSS User's Guide I,II,III. Kaysville: NCSS, 2001

Chajdiak, J: Analýza stavu ekonomiky podnikov a odvetví. Bratislava: STATIS, 1999

Maddala, G. S. : Introduction to Econometrics. London: Prentice-Hall, 1992

Sojka, J.: Matematické modelovanie ekonomických procesov. Bratislava: Alfa, 1986

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

Empirický bayesovský bodový odhad parametra Poissonovho rozdelenia.

Eva Kotlebová

Poissonovo rozdelenie má pri modelovaní stochastických procesov nezastupiteľný význam. Používa sa v situáciách, keď sa sledovaný jav vyskytuje veľmi zriedkavo, pričom počet opakovaní náhodného pokusu je veľmi veľký. Rozdelenie má teda široké uplatnenie nielen v poisťovníctve, ale aj pri modelovaní rôznych prírodných a ekonomických javov.

Poissonovo rozdelenie je špeciálnym prípadom rozdelenia binomického, ktorého parametre *n* (počet opakovaní náhodného pokusu) a π (pravdepodobnosť nastatia udalosti pri realizácii jedného náhodného pokusu) spĺňajú tieto podmienky: $n \to \infty$, $\pi \to 0$ a súčin $n \cdot \pi$ je konštantný. Táto konštanta ($\lambda = n \cdot \pi$) je jediným parametrom Poissonovho rozdelenia. Funkcia hustoty má tvar

$$f(x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$ pre x = 0, 1, 2,..., pričom platí $E(X) = D(X) = \lambda$.

Ak parameter rozdelenia nepoznáme, môžeme ho odhadnúť pomocou údajov získaných náhodným výberom. V klasickej štatistike je najlepším bodovým odhadom λ aritmetický priemer z vybraných údajov.

Bayesovská štatistika rieši problém komplexnejšie. Nielen pri bodových odhadoch neznámych parametrov, ale aj pri iných induktívnych postupoch berie totiž do úvahy nielen údaje z náhodného výberu, ale aj iné informácie o sledovanom parametri, ktoré sú k dispozícii ešte pred realizáciou náhodného výberu. Výsledný odhad je potom kompromisom medzi výsledkom, ktorý sme získali z náhodného výberu, a spracovanou informáciou, ktorú máme k dispozícii okrem údajov z náhodného výberu.

Matematicko-štatistickou podstatou bayesovskej indukcie je nový pohľad na neznámy odhadovaný parameter základného súboru. Oproti klasickej indukcii, v ktorej je považovaný za neznámu konštantu, v bayesovskej indukcii je *náhodnou premennou*, ktorej zákon rozdelenia nám umožní robiť čo najpresnešie závery (bodové a intervalové odhady, testovanie hypotéz).

Pred realizáciou náhodného výberu už teda existuje isté rozdelenie pravdepodobnosti odhadovaného parametra, ktoré sa nazýva *apriórne***.** Po uskutočnení náhodného výberu a jeho spracovaní sa novovzniknuté rozdelenie nazve *aposteriórne*. (Nemusia sa vždy odlišovať – výberový súbor môže potvrdiť platnosť apriórneho rozdelenia.) Vzťah medzi týmito rozdeleniami ukazuje Bayesova formula (uvádzame spojitú verziu):

(1) $$f(\theta / x) = \frac{f(x/\theta).f(\theta)}{\int f(x/\theta).f(\theta)d\theta},$$

kde $f(\theta)$ je apriórne rozdelenie parametra θ, $f(\theta / x)$ je aposteriórne rozdelenie parametra θ, $f(x/\theta)$ je podmienená hustota náhodnej premennej X a $\int f(x/\theta).f(\theta)d\theta$ je tzv. marginálna hustota, ktorá je však z hľadiska premennej θ konštantná. Preto ju pri našich úvahách nebudeme zohľadňovať a budeme používať zápis

(2) $$f(\theta / x) \propto f(x/\theta).f(\theta),$$

kde symbol $\propto$ budeme interpretovať *je proporcionálne* (úmerné). Hustota aposteriórneho rozdelenia parametra θ je teda proporcionálna súčinu hustoty apriórneho rozdelenia a podmienenej hustoty $f(x/\theta)$.

Voľba apriórneho rozdelenia je niekedy veľmi náročná, je však zrejmé, že od neho (najmä v prípade malých výberov) závisí presnosť a spoľahlivosť induktívnych záverov. V niektorých prípadoch sa používajú modely, v ktorých apriórne aj aposteriórne rozdelenia sú rovnakého typu (parametre sa však môžu líšiť). Vtedy hovoríme, že takéto rozdelenie je *konjugované* k rozdeleniu, z ktorého pochádza náhodný výber. Je dokázané, že pre Poissonovo rozdelenie, ktorým sa zaoberá tento príspevok, je vhodným konjugovaným rozdelením Gama rozdelenie. Presnejšie, ak parameter Poissonovho rozdelenia λ má apriórne rozdelenie Gama s parametrami α, β, potom aposteriórne rozdelenie je tiež typu Gama, pričom jeho parametre možno vyjadriť vzťahmi $\alpha' = \alpha + \sum_{i=1}^{n} x_i$, $\beta' = \beta + n$, kde n je rozsah výberového súboru a hodnoty x_i (i = 1, 2, ..., n) pochádzajú z náhodného výberu.

V prípade kvadratickej stratovej funkcie je potom najlepším bayesovským bodovým odhadom stredná hodnota aposteriórneho rozdelenia, t.j.

(3) $$BE(\lambda) = \frac{\alpha'}{\beta'} = \frac{\alpha + \sum_{i=1}^{n} x_i}{\beta + n},$$

kde $BE(\lambda)$ označuje *bayes estimation* (bayesovský odhad parametra λ).

Empirická bayesovská štatistika predstavuje prístup, ktorý je svojou podstatou kdesi medzi klasickým a bayesovským prístupom. Ako apriórnu informáciu využíva predchádzajúce výberové zisťovania, pomocou ktorých možno konštruovať apriórne rozdelenie odhadovaného parametra.

Presnejšie - ak máme k dispozícii k pozorovaní $x_1, x_2, ..., x_k$, na konštrukciu apriórneho rozdelenia parametra θ použijeme zistené hodnoty $x_1, x_2, ..., x_{k-1}$. Toto rozdelenie spolu s údajom x_k s využitím vzťahu (1) umožní vyjadrenie aposteriórneho rozdelenia, ktoré tvorí základ pre nasledujúce induktívne úvahy.

Ak je naším cieľom určenie (len) bodového odhadu neznámeho parametra θ_k, v špeciálnych prípadoch (ak to umožní tvar funkcie hustoty rozdelenia, z ktorého pochádza náhodný výber) nie je

nutné vyjadrovať hustotu apriórneho rozdelenia z pozorovaní $x_1, x_2, ..., x_{k-1}$, následne použitím vzťahu (1) hustotu aposteriórneho rozdelenia - k výsledku možno dospieť aj jednoduchšie. Takýmto prípadom je aj odhad parametra Poissonovho rozdelenia λ v situácii, keď sa použije kvadratická stratová funkcia (t.j. na bodový odhad potrebujeme vedieť len strednú hodnotu aposteriórneho rozdelenia). Podľa (1) možno hustotu aposteriórneho rozdelenia vyjadriť vzťahom

$$q(\lambda / x) = \frac{(e^{-\lambda}\lambda^x / x!) \cdot p(\lambda)}{\int_\lambda (e^{-\lambda}\lambda^x / x!)\, p(\lambda) d\lambda}, \tag{4}$$

kde $p(\lambda)$ označuje apriórnu hustotu (bližšie nešpecifikovanú). Stredná hodnota takto vyjadreného aposteriórneho rozdelenia má tvar

$$E(\lambda_k / x_k) = \frac{\int_\lambda \lambda (e^{-\lambda}\lambda^x / x!) \cdot p(\lambda) d\lambda}{\int_\lambda (e^{-\lambda}\lambda^x / x!)\, p(\lambda) d\lambda}. \tag{5}$$

Označme menovateľa zlomku zo vzťahu (4) zápisom $f(x)$, teda

$$f(x) = \int_\lambda (e^{-\lambda}\lambda^x / x!) \cdot p(\lambda) d\lambda .$$

Potom

$$f(x+1) = \int_\lambda (e^{-\lambda}\lambda^{x+1} / (x+1)!) \cdot p(\lambda) d\lambda ,$$

čo nám umožní vyjadriť vzťah (5) podstatne jednoduchšie:

$$E(\lambda_k / x_k) = (x_k + 1) \cdot \frac{f(x_k + 1)}{f(x_k)}. \tag{6}$$

Ako vidno, v poslednom vzťahu už nevystupuje apriórna (ani aposteriórna) hustota, takže nebolo nutné pokúsiť sa ju analyticky vyjadriť. Zlomok $\frac{f(x_k+1)}{f(x_k)}$ možno pomocou zistených údajov odhadnúť nasledovne:

$$\frac{f(x_k+1)}{f(x_k)} = \frac{počet\ \ vybraných\ \ jednotiek\ \ s\ \ hodnotou\ \ (x_k+1)}{počet\ \ vybraných\ \ jednotiek\ \ s\ \ hodnotou\ \ x_k}, \tag{7}$$

čo predstavuje vlastne podiel frekvencií výskytu dvoch susediacich hodnôt. Takýto odhad má dobré asymptotické vlastnosti, ale v prípade pomerne malého výberu z veľkého základného súboru vedie niekedy k nepresným výsledkom. V nasledujúcom (hypotetickom) príklade ukážeme, ako sa takýto odhad môže líšiť od odhadu získaného klasickým prístupom.

S použitím programov *Statgraphics Plus* a *EXCEL* sme postupne generovali 40 hodnôt apriórneho rozdelenia náhodnej premennej Λ s Gama rozdelením s parametrami $\alpha = 16,\ \beta = 2$ (t.j. $E(\Lambda) = 8,\ \ D(\Lambda) = 4$). Ku každej z nich sme potom generovali hodnoty náhodnej premennej X s Poissonovým rozdelením s príslušnou hodnotou parametra λ. Údaje sme vytriedili a vypočítali sme odhad parametra λ klasicky (priemer z generovaných údajov x_i) aj vyššie popísaným postupom

použitím vzťahu (7). Nasledujúca tabuľka obsahuje generované hodnoty λ_i (2. stĺpec), k nim generované hodnoty x_i (3. stĺpec), ako aj príslušné výpočty.

Tabuľka 1. Generované hodnoty parametra λ a k nim prislúchajúce hodnoty x.

Generované hodnoty		
i	λ_i	x_i
1	4,96059	4
2	7,50457	10
3	6,51422	9
4	5,985	10
5	8,11479	8
6	4,68844	9
7	7,26739	9
8	5,46462	4
9	6,16782	10
10	3,77879	2
11	5,73907	7
12	6,62435	8
13	6,12404	5
14	4,72777	5
15	3,268	1
16	4,60888	2
17	4,75318	3
18	4,3715	4
19	5,69852	3
20	5,45108	3
21	6,6557	6
22	7,65632	6
23	8,90031	7
24	6,08596	4
25	6,52189	2
26	4,56593	4
27	5,58104	8
28	5,56073	8
29	6,06626	4
30	3,45833	3
31	6,36439	2
32	6,19639	7
33	5,68358	2
34	8,79917	11
35	4,96609	2
36	5,80027	6
37	6,92639	8
38	5,09134	2
39	7,85546	7
40	4,56748	3

Vytriedené hodnoty premennej X		
x_i	n_i	$x_i n_i$
1	1	1
2	7	14
3	5	15
4	6	24
5	2	10
6	3	18
7	4	28
8	5	40
9	3	27
10	3	30
11	1	11
spolu	40	218
priemer	**=**	**5,45**

x_{40} =	3
x_{40} + 1=	4
f(3) =	5
f(4) =	6
f(4)/f(3)=	1,2

Podľa vzťahu (7) potom platí $E(\lambda_{40} / x_{40}) = (x_{40} + 1) \cdot \frac{f(x_{40} + 1)}{f(x_{40})} = 4 \cdot \frac{6}{5} = 4{,}8.$

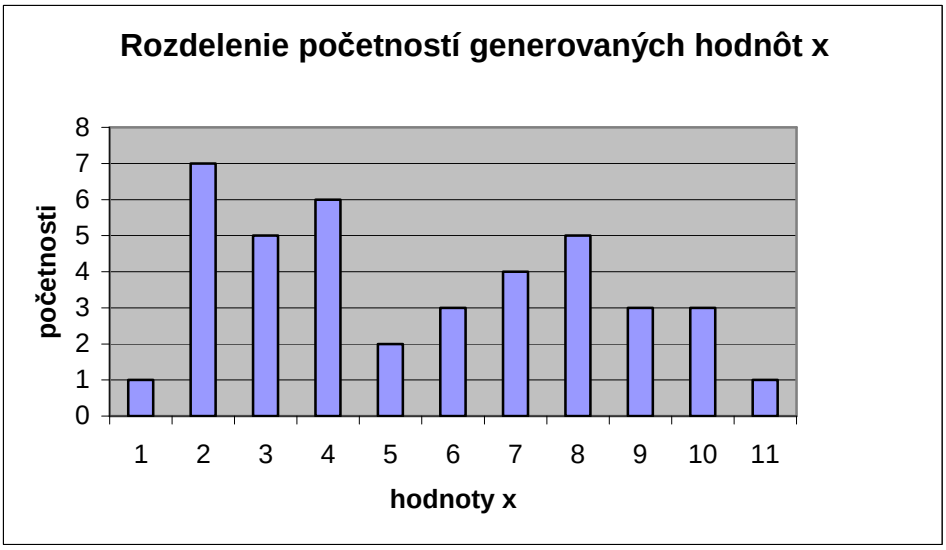
Dostali sme teda takéto výsledky:

klasický odhad: $E(\lambda) = 5{,}45$,

empirický bayesovský odhad (*empirical bayes estimation)*: $EBE(\lambda) = 4{,}8$.

Ako vidíme, získané hodnoty sa dosť odlišujú, čo svedčí o tom, že popísaná metóda, aj keď je pomerne jednoduchá a nevyžaduje si osobitný softvér, nie je univerzálne použiteľná, ale sú potrebné isté predpoklady. Tým najdôležitejším je tvar rozdelenia generovaných hodnôt - malo by ísť o rozdelenie dobre modelovateľné spojitou funkciou, čo v tomto prípade určite nie je splnené, ako vidno z grafu rozdelenia početností:

Obr.1: Histogram rozdelenia početností generovaných hodnôt x_i.



V prípade, že hodnoty premennej X sú modelovateľné spojitým rozdelením, má popísaná metóda veľmi dobré asymptotické vlastnosti a je použiteľná aj pre iné typy rozdelení ako je Poissonovo rozdelenie (napr. geometrické alebo negatívne binomické rozdelenie).

Na záver považujeme za potrebné zdôrazniť, že uvedený príklad bol hypotetický, a jeho poslaním bolo upozorniť na dôležitosť overenia predpokladov ešte predtým, než použijeme metódu, ktorá je jednoduchá a rýchlo nás dovedie ku konkrétnym výsledkom.

Záver:

V článku sme teoreticky porovnali porovnali bodový odhad parametra Poissonovho rozdelenia z pohľadu klasickej štatistiky a prostredníctvom empirického bayesovského prístupu. V druhom prípade sa pristupuje k riešeniu problému komplexnejšie – na náhodný výber môžeme za istých okolností pozerať ako na postupnosť jednoduchých pozorovaní, z ktorých len to posledné považujeme

za náhodne vybranú štatistickú jednotku; tie predchádzajúce použijeme na konštrukciu apriórneho rozdelenia. Ak je toto rozdelenie vytvorené, možno použiť klasické bayesovské metódy na odhad hľadaného parametra. V niektorých prípadoch (ak chceme získať len bodový odhad) sa netreba zaoberať apriórnym rozdelením, existuje nenáročný algoritmus na rýchly výpočet hľadanej hodnoty. Je však nutné, aby údaje, ktoré máme k dispozícii, boli modelovateľné spojitým rozdelením. Ak táto podmienka splnená nie je, empirický bodový odhad môže priniesť veľmi skreslené a zavádzajúce výsledky, ako ukazuje hypotetický príklad prezentovaný v príspevku.

Literatúra:

[1] BERNARDO JOSÉ M. – SMITH ADRIAN F.M.: *Bayesian theory,* Chichester: John Wiley & sons Ltd., 1995

[2] FREUND J.E.: *Mathematical statistics*, New Jersey, Prentice-Hall, International, 1992

[3] GARTHWAITE, P. H. – JOLLIFFE, I.T.: *Statistical Inference*, Prentice-Hall International, Inc., 1995

[4] LEE PETER M: *Bayesian statistics*, New York: Oxford University Press, 1989

[5] PACÁKOVÁ V. a kolektív: *Štatistika pre ekonómov*, Bratislava: IURA EDITION, 2003

[6] PACÁKOVÁ V.: *Aplikovaná poistná štatistika*, Bratislava: IURA EDITION, 2004

[7] WONNACOTT T.H. – WONNACOTT R.J.: *Statistika pro obchod a hospodářství*, Praha: VICTORIA PUBLISHING, 1993

RNDr. Eva Kotlebová, Katedra štatistiky FHI EU v Bratislave

Použitie jednoduchých metód viacrozmerného porovnávania pri analýze životnej úrovne krajov Slovenskej a Českej republiky

Viera Labudová - Zuzana Rybánska

Abstract

The main subject undertaken in presented article is a comparison of living standards in the NUTS3 regions. For the purpose of assessment of living standard of the population in the Czech Republic and in the Slovak Republic we applied a various methods of ordering multidimensional objects.

Obsah pojmu životná úroveň nie je medzi jednotlivými autormi, ktorí sa podrobne zaoberajú touto problematikou, chápaná jednotne. Vymedzenie kategórie životnej úrovne je veľmi zložité a niekedy je predmetom diskusií medzi ekonómami, sociológmi a filozofmi. [3]

Životná úroveň sa obvykle definuje ako súhrn všetkých materiálnych, kultúrnych, sociálnych a morálnych hodnôt, ktoré má obyvateľstvo v danom čase a priestore na uspokojovanie životných potrieb k dispozícii. Pojem „životná úroveň“ je veľmi zložitou, vnútorne členenou kategóriou, ktorá sa skladá z celého radu komponentov.[6]

Životnú úroveň nemožno merať priamo, na jej opis a kvantifikáciu je potrebné použiť celú množinu znakov (vlastností). Jej analýzu možno potom uskutočniť aplikovaním dvoch rôznych prístupov.

Prvý prístup spočíva v budovaní systému ukazovateľov, podrobne charakterizujúcich sledovaný jav. Uvedený systém umožňuje analyzovať jednotlivé prvky zloženého javu a to ustavičným dopĺňaním ukazovateľov, t.j. nie redukciou, ale rozširovaním rozmerov obsiahnutých v danom jave.

Druhý prístup je založený na konštrukcii syntetickej miery (taxonomickej miery rozvoja), ako funkcie premenných meritórne zviazaných na nižších úrovniach agregácie, pričom konštrukcia miery je „spojená s transformáciou viacrozmerného priestoru vybraných znakov do jednorozmerného priestoru agregovanej premennej, ktorej realizácia závisí od všetkých pôvodných premenných.“[7]

Výber a redukcia premenných

Pri výbere premenných, ktoré použijeme na popis životnej úrovne, ak ich hodnotíme na základe ich meritórno-formálnych vlastností, treba zohľadniť nasledovné kritériá: univerzálnosť, merateľnosť, dostupnosť číselných hodnôt, kvalita údajov, ekonomickosť, interpretovateľnosť a spôsob pôsobenia na skúmaný jav.

V súvislosti s posledným, pri výbere premenných uplatňovaným kritériom, môžeme premenné rozdeliť na stimulujúce premenné (stimulanty), destimulujúce premenné (destimulanty) a nominanty.

Stimulanty sú také premenné, ktorých vyššie hodnoty dokazujú vyššiu úroveň rozvoja sledovaného javu.

Destimulantami, naopak, nazývame premenné, ktoré dokazujú vyššiu úroveň rozvoja sledovaného javu svojimi nízkymi hodnotami.

Nominanty sú premenné, ktorých rastúce hodnoty vplývajú pozitívne na sledovaný jav, ale len po určitú hodnotu. Po prekročení tejto hodnoty, nazývanej nominálna hodnota, ich vplyv na sledovaný jav je negatívny.

Premenné, ktoré opisujú zložený jav sú často vzájomne závislé, môžu byť nositeľmi podobných informácií o sledovanom jave. V súvislosti s tým je potrebné pôvodnú množinu premenných redukovať. Pri redukcii sa zohľadňuje objem informácií o sledovanom jave, ktoré obsahujú a vzájomné väzby medzi premennými.

Od premenných, ktoré ostanú po takejto selekcii, sa vo všeobecnosti vyžaduje dostatočná informačná hodnota. To znamená, že nemajú byť podobné v zmysle ich informačného prínosu o sledovanom jave, no súčasne musia obsahovať toľko informácií ako pôvodný, nezredukovaný súbor. Ďalej sa musia vyznačovať dostatočnou priestorovou variabilitou. Vzhľadom na tieto požiadavky treba vstupnú množinu premenných redukovať.

V prvej etape redukcie premenných sa uplatňuje štatistické hľadisko, ktoré je založené na posudzovaní priestorovej variability premenných. Z analýzy sú vylúčené premenné, ktorých variabilita, vyjadrená variačným koeficientom, je veľmi nízka :

$$V_j \le \varepsilon , \qquad (1)$$

$$V_j = \frac{s_j}{\overline{x}_j} , \qquad (2)$$

kde:

ε- arbitrálne zadaná hodnota ($\varepsilon > 0$), najčastejšie $\varepsilon = 0,1$.

V druhej etape sa redukuje pôvodná množina premenných na základe ich informačnej hodnoty. Využívajú sa pritom rôzne miery kvantifikujúce väzby medzi jednotlivými premennými. Medzi najčastejšie využívané miery patrí koeficient korelácie.

Východiskom je korelačná matica premenných vstupujúcich do analýzy:

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{pmatrix} \qquad (3)$$

Za podobné premenné X_i, X_j sa považujú tie, pre ktoré platí:

$$\left| r_{ij} \right| > r^* , \qquad (4)$$

$$\text{kde: } r^* = \min_i \max_j \left| r_{ij} \right| , \qquad (5)$$

pričom r_{ij} *(i , j=1, 2,...,m)* je hodnota párového koeficienta korelácie premenných X_i, X_j.

Konštrukcia syntetickej premennej

Pri tvorbe syntetickej premennej sa stretávame s rôznymi prístupmi, ktoré sa líšia spôsobmi normovania do analýzy vstupujúcich premenných a ich následnej agregácie. V príspevku využívame metódy, ktoré sú v našej štatistickej praxi známe ako jednoduché metódy hodnotenia objektov: metódu súčtu poradí, bodovaciu metódu, metódu normovanej premennej a metódu vzdialenosti od fiktívneho objektu.

Ak sledujeme hodnoty premenných X_j *(j=1,2,...,m) na objektoch* O_i *(i=1,2,...n)*, potom hodnotu x_{ij} považujeme za realizáciu *j-tej* premennej na *i*-tom objekte a jej spôsoby normovania možno vyjadriť pre použité metódy nasledovne:

Metóda súčtu poradí

$$z_{ij} = \begin{Bmatrix} 1 \quad pre \quad \min_i \{ x_{ij} \} \ (j = 1,2,...,m) \\ ..... \\ m \quad pre \quad \max_i \{ x_{ij} \} \end{Bmatrix} , \text{ ak je } X_j \text{ stimulujúca premenná,} \qquad (6)$$

$$z_{ij} = \begin{Bmatrix} 1 \quad pre \quad \max\limits_{i} \{ x_{ij} \} \ (j = 1,2,...,m) \\ ..... \\ m \quad pre \quad \min\limits_{i} \{ x_{ij} \} \end{Bmatrix}$$, ak je X_j destimulujúca premenná. (7)

Bodovacia metóda

$z_{ij} = \frac{x_{ij}}{x_{max,\,j}} \cdot 100$ *(i=1,2,...n; j=1,2,...m),* ak je X_j stimulujúca premenná (8)

$z_{ij} = \frac{x_{\min,\,j}}{x_{ij}} \cdot 100$ *(i=1,2,...n; j=1,2,...m),* ak je X_j destimulujúca premenná (9)

kde:

$x_{max,j}$ - najvyššia hodnota j-tej premennej (ocenená 100 bodmi),
$x_{min,j}$ - najnižšia hodnota j-tej premennej (ocenená 100 bodmi).

Metóda normovanej premennej

$z_{ij} = \frac{x_{ij} - \overline{x_j}}{s_{x_j}}$ *(i=1,2,...n; j=1,2,...m),* ak je X_j stimulujúca premenná (10)

$z_{ij} = \frac{\overline{x_j} - x_{ij}}{s_{x_j}}$ *(i=1,2, ... n; j=1,2, ... m),* ak je X_j destimulujúca premenná (11)

kde:

$\overline{x_j}$ - priemerná hodnota j-tej premennej,
s_{x_j} - štandardná odchýlka j-tej premennej.

Agregovanie normovaných hodnôt bez použitia váh sa pri týchto metódach riadi vzťahom:

$$z_i = \frac{1}{m} \sum_{j=1}^{m} z_{ij} \ \ (i = 1, 2,\ldots, n).$$ (12)

Metóda vzdialenosti od fiktívneho objektu

Normovanie sa uskutočňuje podľa vzťahov (10), resp. (11), syntetická premenná sa konštruuje nasledovne:

$$z_i = \sqrt{\frac{1}{m} \cdot \sum_{j=1}^{m} \left(z_{ij} - z_{0j} \right)^2} \qquad (i=1, 2,..., n).$$ (13)

kde: z_{0j} je normovaná hodnota j-tej súradnice fiktívneho objektu.

Fiktívny objekt predstavuje abstraktný model, ktorý vo všetkých ukazovateľoch dosahuje najlepšie hodnoty súboru (maximálne resp. minimálne podľa charakteru ukazovateľa):

$x_{0j} = x_{\max,\,j}$ pre stimulujúce premenné,
$x_{0j} = x_{\min,\,j}$ pre destimulujúce premenné.

Uvedený spôsob konštrukcie syntetickej premennej uplatňuje princíp rovnakej dôležitosti použitých premenných. Vzťahy pre agregovanie normovaných (transformovaných) hodnôt možno upraviť zakomponovaním váh. Pri ich kvantifikácii možno uplatniť formálno-štatistický prístup, ktorý vychádza z predpokladu, že „dôležitosť danej diagnostickej premennej je proporcionálna jej informačnému obsahu“ [1]. Ak použijeme pri tvorbe váh premenných relatívnu mieru variability, ktorou je variačný koeficient, váhy kvantifikujeme podľa vzťahu:

$$w_j = \frac{V_j}{\sum_{j=1}^{m} Vj} \quad (j = 1, 2, ..., m), \qquad (14)$$

kde:
V_j je variačný koeficient j-tej premennej.

Výsledky analýzy

V tomto príspevku sme sa pokúsili porovnať kraje Slovenskej republiky s krajmi Českej republiky a to z hľadiska dosiahnutej životnej úrovne obyvateľstva. Pri výbere ukazovateľov na analýzu životnej úrovne sa najčastejšie sleduje pokrytie nasledovných oblastí: ochrana zdravia a sociálna ochrana, trh práce, podmienky a bezpečnosť práce, príjmy obyvateľstva, podmienky bývania, vzdelanie, kultúra a voľný čas, dopravná infraštruktúra, ochrana životného prostredia. Náš výber bol ovplyvnený absenciou údajov na strane Slovenskej, resp. Českej republiky a snahou o čo najaktuálnejšie údaje (údaje sú z roku 2003).

Na porovnanie životnej úrovne krajov boli vybrané nasledovné ukazovatele:

- X_1 - Dojčenská úmrtnosť
- X_2 - Počet živonarodených pripadajúcich na 1000 obyvateľov regiónu
- X_3 - Hustota ciest 1. triedy v km/100 km^2
- X_4 - Počet osobných áut pripadajúcich na 1000 obyvateľov regiónu
- X_5 - Počet dokončených bytov v bežnom roku pripadajúcich na 1000 obyvateľov regiónu
- X_6 - Priemerná obytná plocha bytu v m^2
- X_7 - Počet obyvateľov regiónu pripadajúcich na 1 lekára
- X_8 - Počet nemocníc pripadajúcich na 100 tisíc obyvateľov regiónu
- X_9 - Miera nezamestnanosti k 31.12. bežného roka (z disponibilného počtu evidovaných nezamestnaných) – v %
- X_{10} - Počet miest v zariadeniach sociálnych služieb pripadajúcich na 1000 obyvateľov regiónu
- X_{11} - Počet bytových telefónnych staníc pripadajúcich na 1000 obyvateľov regiónu.

Vzhľadom na podmienky, ktoré musia premenné spĺňať, pôvodná množina bola redukovaná, takže do ďalšej analýzy vstupovali premenné X_2, X_3, X_6, X_7, X_9, X_{10}.

Výsledky usporiadania pre premenné, ktorých dôležitosť v danej analýze bola považovaná za rovnakú, uvádza Tabuľka 1. V Tabuľke 2 sú uvedené výsledky analýzy pre premenné, ktorých rôznu dôležitosť sme kvantifikovali váhami (vzťah 14).

Tab.1 *Usporiadanie krajov SR a ČR s využitím nevážených údajov*

Metóda súčtu poradí		Bodovacia metóda		Metóda normovanej premennej		Metóda vzdialenosti od fikt. objektu	
Por.	**Kraj**	**Por.**	**Kraj**	**Por.**	**Kraj**	**Por.**	**Kraj**
1	Morav.-sliezsky	1	Bratislavský	1	Morav.-sliezsky	1	Morav.-sliezsky
2	Bratislavský	2	Hl. mesto Praha	2	Bratislavský	2	Ústecký
3	Zlínsky	3	Morav.-sliezsky	3	Zlínsky	3	Král.-hradecký
4	Král.-hradecký	4	Zlínsky	4	Ústecký	4	Zlínsky
5	Ústecký	5	Plzeňský	5	Plzeňský	5	Pardubický
6	Plzeňský	6	Ústecký	6	Král.-hradecký	6	Bratislavský
7	Pardubický	7	Král.-hradecký	7	Pardubický	7	Liberecký
8	Juhočeský	8	Juhočeský	8	Olomoucký	8	Olomoucký
9	Stredočeský	9	Stredočeský	9	Juhočeský	9	Juhočeský
10-11	Liberecký	10	Pardubický	10	Liberecký	10	Plzeňský
10-11	Olomoucký	11	Olomoucký	11	Stredočeský	11	Žilinský
12	Karlovarský	12	Liberecký	12	Žilinský	12	Stredočeský
13	Žilinský	13	Žilinský	13	Košický	13	Juhomoravský
14	Juhomoravský	14	Košický	14	Hl. mesto Praha	14	Karlovarský
15-16	Košický	15	Juhomoravský	15	Juhomoravský	15	Prešovský
15-16	Hl. mesto Praha	16	Prešovský	16	Prešovský	16	Vysočina
17	Vysočina	17	Karlovarský	17	Karlovarský	17	Košický
18	Prešovský	18	Nitriansky	18	Vysočina	18	Trnavský
19	Nitriansky	19	Trnavský	19	Trnavský	19	Nitriansky
20	Trnavský	20	Banskobystrický	20	Nitriansky	20	Banskobystrický
21	Banskobystrický	21	Vysočina	21	Banskobystrický	21	Hl. mesto Praha
22	Trenčiansky	22	Trenčiansky	22	Trenčiansky	22	Trenčiansky

Zdroj: vlastné výpočty

Tab. 2 *Usporiadanie krajov SR a ČR s využitím vážených údajov*

Metóda súčtu poradí		Bodovacia metóda		Metóda normovanej premennej		Metóda vzdialenosti od fikt. objektu	
Por.	**Kraj**	**Por.**	**Kraj**	**Por.**	**Kraj**	**Por.**	**Kraj**
1	Bratislavský	1	Bratislavský	1	Bratislavský	1	Král.-hradecký
2	Zlínsky	2	Hl. mesto Praha	2	Zlínsky	2	Zlínsky
3	Král.-hradecký	3	Zlínsky	3	Morav.-sliezsky	3	Morav.-sliezsky
4	Plzeňský	4	Plzeňský	4	Plzeňský	4	Bratislavský
5	Juhočeský	5	Morav.-sliezsky	5	Král.-hradecký	5	Pardubický
6	Morav.-sliezsky	6	Juhočeský	6	Hl. mesto Praha	6	Liberecký
7	Pardubický	7	Král.-hradecký	7	Juhočeský	7	Juhočeský
8	Stredočeský	8	Stredočeský	8	Pardubický	8	Plzeňský
9	Hl. mesto Praha	9	Pardubický	9	Stredočeský	9	Olomoucký
10	Liberecký	10	Liberecký	10	Liberecký	10	Ústecký
11	Ústecký	11	Ústecký	11	Olomoucký	11	Stredočeský
12	Olomoucký	12	Olomoucký	12	Ústecký	12	Juhomoravský
13	Karlovarský	13	Žilinský	13	Juhomoravský	13	Žilinský
14	Juhomoravský	14	Juhomoravský	14	Žilinský	14	Karlovarský
15	Vysočina	15	Trenčiansky	15	Karlovarský	15	Vysočina

16	Žilinský	16	Trnavský	16	Vysočina	16	Hl. mesto Praha
17	Trnavský	17	Vysočina	17	Trnavský	17	Trnavský
18	Trenčiansky	18	Karlovarský	18	Trenčiansky	18	Trenčiansky
19	Nitriansky	19	Nitriansky	19	Nitriansky	19	Nitriansky
20	Košický	20	Košický	20	Prešovský	20	Prešovský
21	Prešovský	21	Prešovský	21	Košický	21	Košický
22	Banskobystrický	22	Banskobystrický	22	Banskobystrický	22	Banskobystrický

Zdroj: Vlastné výpočty

Výsledky usporiadania naznačujú veľmi dobré umiestnenie Bratislavského kraja, ktorý vykazuje najnižšiu *mieru nezamestnanosti* a druhú najväčšiu *priemernú obytnú plochu bytu.* Práve týmto premenným pripisujú veľkú dôležitosť použité váhy. Moravsko-sliezsky a Zlínsky kraj sa umiestňovali zvyčajne na prvých troch miestach a z hľadiska vybraných ukazovateľov dosahovali najlepšiu životnú úroveň spomedzi krajov ČR. Hlavné mesto Praha dosiahlo najlepšie výsledky pri použití bodovacej metódy, kde obsadilo druhé miesto. Pri použití ostatných metód sa umiestňovalo v strede tabuliek poradí alebo až na ich konci. Najhoršie výsledky sme u tohto kraja zaznamenali pri použití metódy vzdialenosti od fiktívneho objektu. Tento fakt sa dá pripísať aj tomu, že napriek všeobecne známej vyspelosti spomenutého kraja, pri nami vybraných ukazovateľoch dosahuje najlepšie výsledky iba u dvoch premenných.

Kraje ČR, s výnimkou Karlovarského kraja a kraja Vysočina, sa zväčša nachádzali v prvej polovici tabuliek poradí a nechávali posledné miesta pre šesť krajov SR. Jedine Žilinský kraj sa ani raz neobjavil v skupine posledných krajov. V priemere dosiahol dvanáste miesto, ak uvažujeme len metódy jednoduchého viacrozmerného porovnávania bez použitia váh. Dôvodom zaostávania našich šiestich krajov za krajmi ČR, vzhľadom na vybrané ukazovatele, je vysoká *miera nezamestnanosti*, vysoký *počet obyvateľov na 1 lekára* a nízky *počet miest v zariadeniach sociálnych služieb*. Na posledných troch miestach v tabuľkách porovnaní sa zvyčajne umiestňovali Trenčiansky, Nitriansky a Banskobystrický kraj. Banskobystrický kraj je pri použití každej metódy s využitím váh na poslednom mieste, čo je spôsobené najvyššou *mierou nezamestnanosti*, ktorej pripisujú váhy veľkú dôležitosť.

Literatúra:

[1] ABRAHAMOWICZ, M. - ZAJAC, K. *Metoda wazenia zmiennych w taksonomii numerycznej* ..., 1986.

[2] BARTOSZEWICZ, S. 1976. Propozycja metody tworzenia zmiennych syntetycznych. In: *Prace naukowe AE we Wroclawiu.* Wroclaw, 1976. nr.84.

[3] KEJKULA, J.: Možné přístupy porovnání zemí podle dosaženého stupně životní úrovně. In: *Statistika*, č. 4, 1981, str. 122 – 136

[4] KŘOVÁK, J.: *Vybrané metody víceaspektního hodnocení průmyslových organizací*. VÚSEI, Praha 1983.

[5] LIPIETA, A. et al. *Taksonomiczna analiza przestrzennego zróznicowania poziomu zycia w Polsce w ujeciu dynamicznym*, 2000.

[6] MORAVOVÁ, J. et.al..: *Sociálněhospodářska statistika II*. VŠE, Fakulta informatiky a statistiky, Praha 1997.

[7] Sylaby, T. 1994. *Systemy vskaźnikow spolecznych w polskich warunkach transformacji rynkowej: Monografie I opracowania*. Warszawa: SGH , 1994, nr. 392.

Kontakt: RNDr. Viera Labudová, FHI EU v Bratislave, labudova@dec.euba.sk
Ing. Zuzana Rybanská, La Quinta Inn ,Williamsburgu, Virginia USA, zuzana_rybanska@hotmail.com

Meranie chudoby v krajinách Európskej únie a na Slovensku

Erika Ľapinová

Abstract
Different concepts of poverty. Poverty measuring in EU countries. Common indicators of poverty and social exclusion in EU, national indicators of poverty in Slovak republic.

Cieľom môjho príspevku bude v úvode naplniť aspoň čiastočne pojem chudoba obsahom, následne predstavím spoločné indikátory chudoby a sociálneho vylúčenia pre krajiny Európskej únie a zdroje dát, ktoré sa pri ich vykazovaní používajú. Vzhľadom na komplexnosť problematiky merania chudoby a sociálneho vylúčenia si nekladiem za cieľ postihnúť ju v celom rozsahu. Dovoľujem si dať do pozornosti čitateľa, ktorého téma zaujme, zdroje literatúry, z ktorých som vychádzala pri spracovaní príspevku.

1. Koncepty chudoby

Podľa sociológa prof. PhDr. Petra Ondrejkoviča, PhD., za najvhodnejšie východisko možno pokladať tieto prístupy k definícii chudoby:

A) Tzv. absolútna koncepcia chudoby vychádza z nedosahovania stanoveného absolútneho životného minima, pričom sa vychádza zo stanovených všeobecne uznávaných štandardov bývania, stravovania, obliekania. V SR sa stretávame s pojmom hmotná núdza, ktorý je príkladom pokusu o objektivizáciu chudoby ako sociálneho javu. Znamená stav pod úrovňou životného minima.

B) V duchu definície relatívnej chudoby možno za chudobného považovať každého, kto nedosahuje relatívne štandardy, závislé od bohatstva spoločnosti. T. j. chudobný v bohatšej krajine nebude porovnateľný s chudobným v hospodársky menej prosperujúcej krajine. Najčastejšie sa vychádza z určitého percenta príjmov na obyvateľa. Pojem relatívnej chudoby býva dopĺňaný životnou situáciou (vybavením domácnosti, vzdelaním, sociálnymi kontaktmi, zdravím a i.).

C) Podľa subjektívnej koncepcie chudoby je chudobný ten, kto má podľa subjektívneho odhadu tak málo prostriedkov na živobytie, že mu nedostačujú na základné životné potreby. Pritom samotný pojem základných životných potrieb je neobyčajne variabilný, v závislosti na tradíciách, kultúrnom prostredí, geografických a klimatických podmienkach, spôsobe života indivídua a pod.

D) Podľa politickej koncepcie chudoby je chudobný ten, kto spĺňa kritériá stanovené štátom na poberanie sociálnych dávok, resp. sociálnej pomoci v chudobe. V tomto prípade sú počty chudobných v príslušnej krajine, regióne, závislé od počtu poberateľov uvedenej pomoci. Politická koncepcia chudoby vychádza z nevyhnutnosti sporadického prispôsobovania kritérií a hraníc chudoby možnostiam štátu, regiónu. Tým sa blíži ku koncepcii relatívnej chudoby.

Ako uvádza Roman Džambazovič, z Katedry sociológie Filozofickej fakulty Univerzity Komenského v Bratislave a výskumný pracovník Strediska pre štúdium práce a rodiny v Bratislave, hoci sa pri skúmaní chudoby postupne upúšťa od je jednodimenzionálneho, značne zjednodušujúceho vymedzenia a merania, výška dosiahnutého peňažného príjmu zostáva aj v súčasnosti základným ukazovateľom chudoby na Slovensku (peňažná, monetárna chudoba). Súvisí to napríklad s pomerne jednoduchým spracovaním, jednoduchšou operacionalizáciou, možnosťou komparácie. Chudoba sa ešte vždy vymedzuje

na základe (nedostatočnosti) príjmu či spotreby a ostatné aspekty chudoby nie sú v diskusii o nej medzi tvorcami sociálnej politiky akoby akceptované.

2. Otvorená metóda koordinácie politík krajín EÚ v oblasti sociálnej inklúzie

Jedným z cieľov Lisabonskej stratégie Európskej únie je boj proti chudobe a sociálnej exklúzii. Európska rada v Lisabone v roku 2000 sa uzniesla na tom , že rozsah chudoby a sociálnej exklúzie je neakceptovateľný a je nevyhnutné podniknúť kroky smerom k podstatnému zníženiu chudoby do roku 2010. To si vyžiadalo rozšírenie nástrojov koordinácie politických opatrení členských štátov zavedením otvorenej metódy koordinácie (OMC).

Medzi kľúčové súčasti Otvorenej metódy koordinácie v oblasti sociálnej inklúzie patria okrem iného i spoločné indikátory na meranie chudoby a sociálnej exklúzie, aby bolo možné monitorovať pokrok prijatých opatrení a porovnávať dobré skúsenosti. Poslednou z ôsmich konkrétnych výziev v rámci Spoločného memoranda o inklúzii (JIM), ktoré podpísala Slovenská republika, je rozšírenie štatistických systémov a indikátorov o chudobe a exklúzii. Obsahuje tieto čiastkové ciele:

- Zosúladiť štatistické systémy a indikátory tak, aby boli kompatibilné s Eurostatom.
- Zabezpečiť viac informácií o rizikových skupinách ako sú Rómovia, ex-väzni, bezdomovci, ľudia so zdravotným postihnutím.
- Zaviesť rod ako povinnú triediacu kategóriu do štatistických systémov a indikátorov v oblasti chudoby a exklúzie.

Národný akčný plán inklúzie implementuje spoločné ciele EÚ v oblasti chudoby a sociálnej exklúzie do národných cieľov, opatrení a programov, berúc do úvahy priority, na ktorých sa dohodli vláda SR a Európska komisia už v Spoločnom memorande o inklúzii. V NAP/inklúzie sú po prvýkrát prezentované údaje o chudobe na základe spoločných indikátorov, čo umožňuje ich komparáciu s ostatnými krajinami EÚ. Európska únia používa tieto definície chudoby, sociálnej exklúzie a sociálnej inklúzie:

Chudoba

Ľudia žijú v chudobe, ak ich príjem a iné zdroje sú natoľko nedostatočné, že im neumožňujú dosiahnuť takú životnú úroveň, ktorá je akceptovateľná v spoločnosti, v ktorej žijú. V dôsledku chudoby môžu poznať mnohonásobné znevýhodnenie od nezamestnanosti, cez nízky príjem, zlé bývanie, nedostatočnú zdravotnú starostlivosť až po prekážky v prístupe k celoživotnému vzdelávaniu, kultúre, športu či rekreácii. Sú často marginalizovaní a vylúčení z účasti na aktivitách (ekonomických, sociálnych a kultúrnych), ktoré sú bežné pre ostatných ľudí a ich prístup k základným právam môže byť obmedzený.

Sociálna exklúzia

Sociálna exklúzia je proces, ktorého prostredníctvom sú určití jednotlivci vytláčaní na okraj spoločnosti a je im zabránené plne na nej participovať v dôsledku svojej chudoby, nedostatku základných kompetencií (spôsobilosť) a príležitostí celoživotného vzdelávania alebo v dôsledku diskriminácie. Toto ich vzďaľuje, izoluje od zamestnania, príjmu a príležitostí vzdelávania, ako aj sociálnych a komunitných sietí a aktivít. Majú veľmi obmedzený prístup k rozhodovacím orgánom, a tak často pociťujú bezmocnosť a nemožnosť riadiť a kontrolovať rozhodnutia, ktoré majú dosah na ich každodenný život.

Sociálna inklúzia

Sociálna inklúzia je proces, ktorý zabezpečuje, aby tí, ktorí sú v riziku chudoby a sociálnej exlúzie, získali príležitosti a nevyhnutné zdroje na to, aby mohli plne participovať

na ekonomickom, sociálnom a kultúrnom živote a mali takú životnú úroveň a blahobyt, ktorý je považovaný za obvyklý v spoločnosti, v ktorej žijú. Zabezpečuje im väčšiu účasť na rozhodovaní, čo ovplyvňuje ich životy a prístup k základným právam.

Miera rizika chudoby

Podiel osôb s ekvivalentným disponibilným príjmom pod 60%-tami mediánu národného ekvivalentného príjmu. Ekvivalentný príjem je definovaný ako podiel celkového disponibilného príjmu domácnosti a jej „ekvivalentného počtu členov“, čím sa berie do úvahy veľkosť a zloženie domácnosti a je prisúdený každému členovi domácnosti. Ekvivalentný počet členov domácnosti je vypočítaný podľa ekvivalentnej škály (modified OECD scale), ktorá priraďuje jednotlivým členom domácnosti nasledovnú váhu: prvému dospelému členovi domácnosti váhu 1; ostatným dospelým členom váhu 0,5; deťom do 14 rokov váhu 0,3 a deťom 14 a viac ročným váhu 0,5.

3. Meranie príjmovej chudoby
Laekenské indikátory založené na príjme
a ďalšie spoločné indikátory chudoby a sociálne exklúzie v krajinách EÚ
Národné indikátory chudoby

Ani názory na meranie príjmovej chudoby nie sú jednoznačné. Kým Európska únia sleduje relatívnu chudobu (indikátory miery rizika chudoby sú postavené na koncepte relatívnej chudoby), v Správe o miléniových rozvojových cieľoch Rozvojového programu Spojených národov (United Nations Development Programme, UNDP) sa hovorí o extrémnej chudobe. Ide o situácie, keď jednotlivec alebo domácnosť nie sú schopné uspokojovať základné ľudské potreby. Podľa tejto správy je zaužívanou hranicou chudoby pre krajiny na porovnateľnom stupni rozvoja ako Slovensko príjem na úrovni 2,15 dolára na osobu a deň (2,15 PPP$).

Laekenské indikátory

Všetky primárne i sekundárne indikátory chudoby a sociálnej exklúzie a ich definície možno nájsť na webovských stránkach Európskej únie, na informačných portáloch venovaných EÚ aj na stránke Ministerstva práce, sociálnych vecí a rodiny Slovenskej republiky.

Rovnako ako spoločné ciele Európy v boji proti chudobe a sociálnemu vylúčeniu, aj indikátory zachytávajú mnohodimenzionálnosť chudoby a sociálnej exklúzie. Takmer polovica z nich je konštruovaná na báze príjmov, tzv. monetárne indikátory, ďalšia polovica má zachytávať ostatné dimenzie, ktorými je prístup k zamestnaniu, vzdelaniu, zdravotnej starostlivosti, bývaniu. Väčšina indikátorov je založená na báze relatívnej chudoby.

PhDr. Zuzana Kusá, CSc., vedecká pracovníčka Sociologického ústavu Slovenskej akadémie vied, na každoročne organizovanom odbornom seminári venovanom otázkam chudoby zdôrazňuje: „ Meranie chudoby je politická otázka. Uprednostnenie tej či onej definície a metódy merania chudoby má podľa nej politickú potenciu a môže mať politické dôsledky. Boli to tzv. laekenské indikátory, ktoré pristupujú k chudobe ako k príjmovej nerovnosti. Meranie chudoby ako príjmovej nerovnosti je politicky značne zraniteľné. Ak totiž chápeme chudobu ako nerovnosť a sústredíme sa na jej relatívne miery a zisťovanie, koľkokrát majú jedni viac alebo menej ako druhí, zbavujeme problém chudoby jeho pálčivosti. Menej naliehavé je to, čo v relatívnej podobe bude existovať vždy, pretože vždy bude mať niekto relatívne menšie príjmy.“

4. Na margo národnej hranice chudoby

Slovensku zatiaľ chýba sústavný výskum národnej hranice chudoby. Ing. Silvia Rybárová, CSc., z Odboru stratégie sociálnej politiky Ministerstva práce, sociálnych vecí a rodiny, ktorá sa podieľala na vypracovaní Spoločného memoranda o inklúzii (JIM) a Národného akčného plánu sociálnej inklúzie 2004 – 2006 (NAP/Inklúzie), uzatvára svoj príspevok na tému Spoločné indikátory chudoby a sociálnej exklúzie konštatovaním potreby národných indikátorov chudoby. Do roku 2001 sa za národnú hranicu chudoby na Slovensku považovalo životné minimum a podiel obyvateľstva pod touto hranicou za indikátor chudoby. V súčasnosti pomoc v hmotnej núdzi už nedosahuje výšku životného minima, počet poberateľov dávky pomoci v hmotnej núdzi sa nekryje s počtom osôb nachádzajúcich sa pod hranicou životného minima. Sociálne štatistiky Ministerstva práce, sociálnych vecí a rodiny SR už adekvátne neodrážajú chudobu na Slovensku. Vytvorenie špecifického národného indikátora chudoby je dôležité aj vzhľadom na to, že spoločne schválená relatívna hranica chudoby na úrovni 60% mediánu ekvivalentného príjmu je veľmi ťažko zdolateľná v súčasnej sociálnoekonomickej situácii. Na základe tejto hranice možno len veľmi ťažko zaznamenávať každoročný pokrok v znižovaní chudoby. Národné indikátory je potrebné špecifikovať aj na meranie takých rizikových skupín obyvateľstva, ktoré nezachytí ani štatistické zisťovanie EU SILC. Autorka príspevku konštatuje: „V Slovenskej republike chýbajú takéto indikátory, a predovšetkým, chýba permanentný výskum, ktorý by umožnil neustále sledovanie dosahovaného pokroku v oblasti sociálnej inklúzie.“ Hovorí o dvoch cestách merania chudoby a sociálnej exklúzie na Slovensku:

1. rozvíjanie spoločných indikátorov v rámci Európskej únie na báze pripravovaného štatistického zisťovania EU SILC, realizovaného Štatistickým úradom SR.
2. vytvorenie škály národných monetárnych i nemonetárnych indikátorov založených na permanentnom výskume istých špecifík SR (osobitne meraní regionálnych rozdielov, takých marginalizovaných skupín ako sú rómske komunity, bezdomovci a ďalšie zraniteľné skupiny, s akceptovaním rodovej dimenzie), ktoré by vhodne doplnili spoločné indikátory a umožnili lepšie špecifikovať chudobu a sociálnu exklúziu na Slovensku.

Realizovať túto úlohu je možné len pomocou spolupráce širokej odbornej verejnosti, vrátane tých, ktorých sa chudoba a sociálna exklúzia bytostne dotýka. Tu vidí veľkú potrebu rozprúdiť širšiu diskusiu o meraní chudoby a sociálnej inklúzie, najmä na národnej úrovni.

Súčasné národné štatistické zdroje monitorujúce sociálnu situáciu slovenskej populácie (Rodinné účty, Mikrocenzus, Výberové zisťovanie pracovných síl), plne nepokrývajú požiadavky Eurostatu na monitorovanie chudoby a sociálnej exklúzie. Z tohto dôvodu sa Štatistický úrad Slovenskej republiky rozhodol na základe odporučení Eurostatu zapojiť sa do prípravy realizovania nového harmonizovaného štatistického nástroja – Štatistiky príjmov a životných podmienok (EU-SILC).

Úlohou zisťovania EU-SILC bude získanie reprezentatívneho zdroja porovnávacích ukazovateľov v oblasti príjmového rozdelenia a sociálnej vylúčenosti. Výsledkom budú dve databázy s indikátormi v prierezovom a dlhodobom pohľade, ktoré poskytnú informácie za domácnosti aj jednotlivcov. Oblasti zisťovania sa budú týkať predovšetkým: príjmov, zdravia, vzdelania, práce, bývania. Realizácia zisťovania sa predpokladá do roku 2006.

Záver

Na koniec si dovolím citovať Romana Džambazoviča a záver jeho príspevku Posun od merania chudoby k meraniu sociálneho vylúčenia: „Indikátory sociálneho vyčlenenia by mali spĺňať nasledujúce kritériá: jednoduchá pochopiteľnosť pre verejnosť, relatívne jednoduchá kvantifikovateľnosť, berúce do úvahy medzinárodné konvencie – možnosť komparácie,

majúce na pamäti dynamickú dimenziu – snaha o zachytenie procesu vylúčenia. Mali by byť vhodné pre operacionalizáciu i na úrovni lokálneho priestoru, majú mať jasnú a akceptovanú normatívnu interpretáciu, majú byť vhodné na politickú intervenciu, ale nie prostriedkom na manipuláciu a v neposlednom rade by mali byť dočasné – čiže vhodné pre revíziu."

Použitá literatúra

[1] DŽAMBAZOVIČ, R. Posun od merania chudoby k meraniu sociálneho vyčlenenia. In: Zborník Otázky merania chudoby. 2004. Bratislava, Friedrich Ebert Stiftung. ISBN 80-89149-02-2.
[2] MAREŠ, P. Chudoba v České republice v datech (šetření sociální situace domácností). Dílčí studie projektu o možnostech monitorování chudoby v ČR. Výzkumný ústav práce a sociálních věcí Praha, Výzkumné centrum Brno, 2004.
[3] ONDREJKOVIČ, P. Chudoba – spoločensky nežiaduci jav. In: Zborník Otázky merania chudoby. 2004. Bratislava, Friedrich Ebert Stiftung. ISBN 80-89149-02-2.
[4] RYBÁROVÁ, S. Spoločné indikátory chudoby a sociálnej exklúzie. In: Zborník Otázky merania chudoby. 2004. Bratislava, Friedrich Ebert Stiftung. ISBN 80-89149-02-2.
[5] SÚKENÍKOVÁ, H. Štatistické monitorovanie chudoby. In: Sociálna pomoc a motivácia k práci. Zborník z konferencie Od závislosti k práci. Bratislava: Výskumný ústav práce, sociálnych vecí a rodiny, júl 2003. ISBN: 80-7138-116-0.
[6] Spoločné memorandum o inklúzii, Slovenská republika, Ministerstvo práce, sociálnych vecí a rodiny SR, Bratislava 2003.
[7] Národný akčný plán sociálnej inklúzie 2004 – 2006 (NAP/inklúzie), Slovenská republika, Ministerstvo práce, sociálnych vecí a rodiny SR, Bratislava, júl 2004.
[8] Zborník Otázky merania chudoby. 2004. Bratislava, Friedrich Ebert Stiftung. ISBN 80-89149-02-2.
[9] Zisťovanie o príjmoch a životných podmienkach domácností EU-SILC 2005 (www.statistics.sk).
[10] Miléniové rozvojové ciele. Cesta k znižovaniu chudoby a sociálneho vylúčenia. Slovenská republika. Správa o miléniových rozvojových cieľoch. Bratislava: Centrum pre hospodársky rozvoj. United Nations Development Programme, 2004. ISBN 92-95042-01-8.
[11] http://www.euractiv.sk/cl/43/4754/Socialny_system_a chudoba
[12] http://www.euroinfo.gov.sk
[13] Je vo Veľkej Británii väčšia chudoba ako v Bulharsku? In: Bulletin Inštitútu sociálnej politiky č. 3/04. Ministerstvo práce, sociálnych vecí a rodiny Slovenskej republiky, Inštitút sociálnej politiky.
[14]http://europa.eu.int/comm/employment_social/social_protection_commitee
[15]http://epp.eurostat.cec.eu.int/

Ing. Erika Ľapinová
Oddelenie ekonomických vied a vied o riadení
Ústavu vedy a výskumu Univerzity Mateja Bela v Banskej Bystrici
Cesta na amfiteáter 1
974 01 Banská Bystrica
Tel.: 048/446 6234, e-mail: erika.lapinova@umb.sk.

DEMOGRAPHICAL AND ECONOMICAL INDICATORS OF THE COUNTRIES WITH LOW SCALE OF ECONOMICAL DEVELOPMENT

Linda Bohdan, Kubanová Jana

Abstract

Majority of articles dealing with analysis of economical and demographical indicators is devoted to the European countries. This paper is on the other hand devoted to one country lying far from Europe, which is on the low level of development and life in many parts of this country reminds the Middle Ages sporadically mixed with advanced elements of technical development.

One of the countries in Asia that can be included in the group of the countries with low scale of economical development is Kingdom of Nepal. Nepal is considered to be one of the richest countries in the world in terms of bio-diversity due to its unique geographical position and altitude variation. Altitude of the country changes from 60 meters above sea level to the 8 848 meters. The highest point on earth, Mt. Everest, lies in the northern part of the country.

Economic process

It is possible to characterise Nepal as a developing country with an agricultural economy. The country grasped to expand into manufacturing industries and other technological sectors. Much progress was noted in recent years. The main economic activities followed by manufacturing are farming, trade and tourism.

Nepal noted economic growth from election of democratic government in the year 1991. Many reforms were initiated, which aimed at economies opening and tendency of them were to accelerate development of the country and to improve economic-social status of population. Growth between 6 - 7% was registered for a few years in non-agricultural sector, economic transformation proceed successfully in other spheres. Political instability in half of 90 years unfortunately slowed down the economic development.

There is substantial impact of political situation to results of economies in Nepal. Growth of GDP stopped in 2002 and economics recorded drop about 0,6 %. Aggravation of political situation influenced tourism and manufacturer sector. Economic growth was restored in years 2002/03 but only about 2,3 %, which is absolutely insufficient level. Production in agricultural sector increased at least (2,4 %). Certain chance to increase interest of tourists was 50th anniversary of the first climbing on the Mount Everest but total number of tourist is however smaller than in the year 1999.

Nepal gets high foreign aid; however fundamental problem is bad distribution of this aid that is caused by functionless administration and by corruption in public administration. Serious problem of Nepal economies is very low level of deposits and investments. Application of economic reforms trends above all to liberalization of rules that are concerned with foreign investments, to further privatization of state enterprises and to reform of financial sector.

Industry

Industrial sector isn't significant factor at GDP formation. It is influenced by little home market, by various administrative measured, by imports from India and geographical location country (it hasn't access to sea). Additional role plays insufficient infrastructure, lack

of qualified workers and last but not least lack of capital. During 2001/02 this sector contributed to the GDP formation only by 8,1 %. Production is above all concentrated into 3 branches, in which is added value very low – production food and beverage, tobacco and textile. Only one sector, that produces allowance from export, is textile industry namely production of ready-made clothing and woollen carpets. The other industries are leather products, paper and cement, steel utensils, cigarettes and sugar.
Investments into industry development are very low, home investors prefer traditional sources like is gold, gemstones and soil. The biggest investor is India with its 170 joint ventures. Most of Nepal's industries are based in the Kathmandu Valley and a string of small towns in the southern Terai plains.

Agriculture

Agricultural sector is in the contrary to industry decisive item of GDP formation. 65 % of active population is employed in agriculture and this sector contributes almost by 40 % to GDP formation. Labour productivity has low level. Agriculture suffers mainly by unpredictable weather. Majority of agricultural plane is in mountainous landscape and retarded way of cultivation conforms to it. Irrigation is also ensured insufficiently. Feudal system landownership enforces all the time, which is characteristic by concentrated ownership and by very high lease charges. Even if down located areas produce enough cereals, this production cannot cover needs of inhabitants in mountain regions and much of food must be imported. Primarily rice is produced (57 % from whole), maize (21%) and corn (17%), further potatoes and tea. Rice is the staple diet in Nepal and around three million tons are produced annually. Besides food grains, cash crops like sugarcane, oil seeds, tobacco, jute are also cultivated in large quantities.
Rolling fields and neat terraces can be seen all over the Terai flatlands and the hills of Nepal. Even in the highly urbanized Kathmandu Valley, large tracts of land outside the city areas are devoted to farming.

Trade

Trade has been a major occupation in Nepal since early times. Nepal is situated at the crossroads of the ancient trans-Himalayan trade route so it is possible to say that trading is second nature to the Nepali people. Foreign trade is characterized mainly by import of manufactured products and export of agricultural raw materials. Nepal imports manufactured goods and petroleum products worth about US$ 1 billion annually. The value of exports is about US$ 315 million. Nepal's largest export articles are carpets, earning the country over US$ 135 million per year. Garment exports account for more than US$ 74 million and handicraft goods bring in about US$ 1 million. Other important exports are pulses, hides and skins, jute and medicinal herbs.

Services

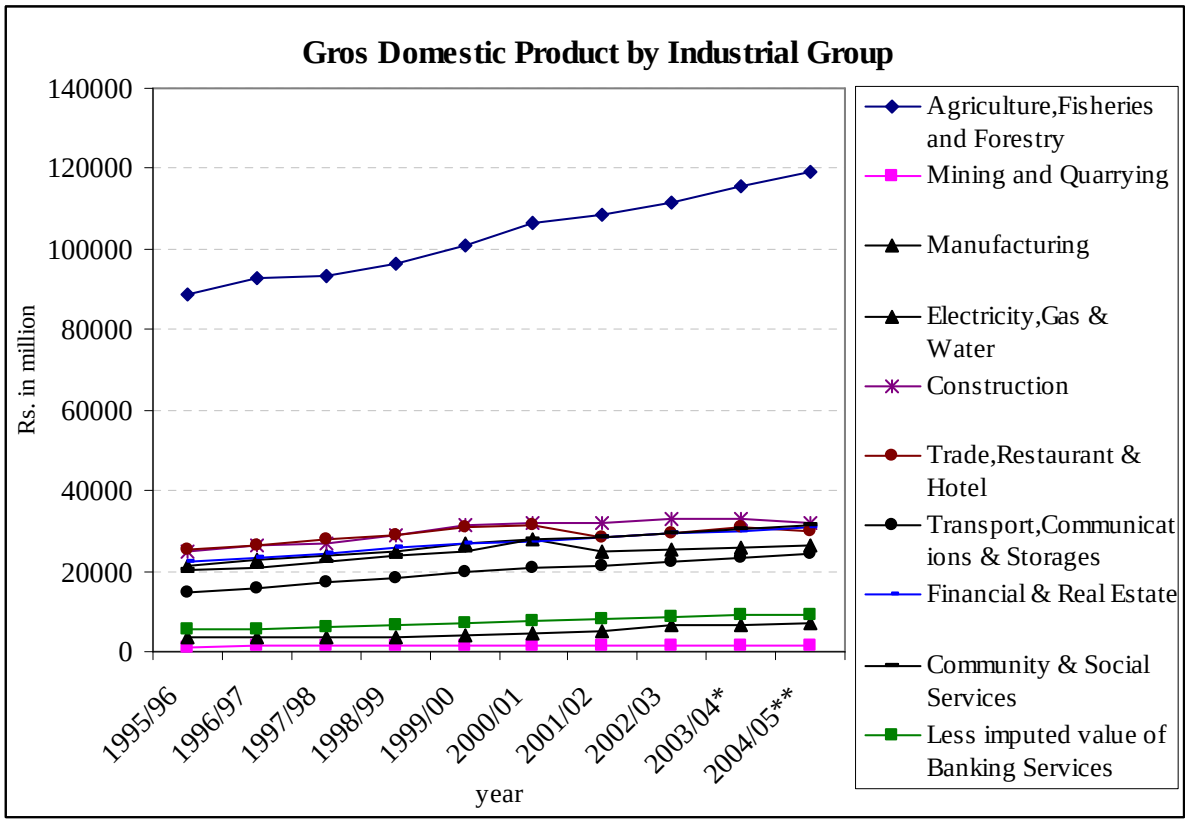
Exceptional positions in sphere services in Nepal occupy tourism. Till the year 1999 the number of tourist visiting Nepal regularly increased about more than 10 % per year to almost half a million visitors. Inflow of tourist decreases to 365 thousands visitors in the year 2001 and 216 thousands in the year 2002. In consequence of this many hotels had to be closed and their holders require governments about help at paying-off borrowings. The number of tourist increased to 265 thousands in the year 2003 but it is about one half of the tourists in 1999. The Nepal government arranged advertising campaign commemorating 50th anniversary of the first climb on Mount Everest what helped to increasing the number of tourists. Main campaign was „ Destination Nepal 2002-2003" for journalist, travel agencies and potential visitor, which was targeted on tourist from India, but also from Western Europe.

Aims of it were to attract half a million of visitors till the end of the year 2003 and to ensure income from tourism about 180 millions USD. The government enabled climbing to other 13 peaks and simplified administration concerning charges for climbing on the peaks.

It is necessary to improve infrastructure for the next increasing of earnings from tourism. Development of tourist routes, which are more available to foreign visitor, is essential in this respect. Environmental science is also important.

Basic macroeconomic indicators (end of financial year is 15.7.)

	1998/99	1999/00	2000/01	2001/02	2002/03
GDP (mld. NPR)	342,0	378,0	410,2	420,2	446,18
Inter-annual growth of GDP (in %)	4,5	6,4	5,8	-0,6	2,3
GDP per inhabitant (in USD)	210	220	237	227	236
Inflation rate	8,1	1,5	6,5	2,8	4,8
Balance of checking account (in mil. USD)	- 256,5	- 277,4	- 339,3	- 254,2	- 155,9
Development of exchange rate of NPR /USD	68,2	71,1	74,9	77,8	74,51
interest rate of commercial banks	14,0	11,33	9,46	7,67	8,0



Demographic data

year	population (in millions)	population density (per 1 km^2)	Economically active population (in millions)	annual increment
1961	9,413	65	4,61	-

1971	11,556	80	5,47	2,07%
1981	15,023	102	6,911	2,66%
1991	18,491	126	8,69	2,1%
1995	20,7	141	9,73	2,1%
2001*	22,7	167	11,2	2,3%

*census 2001

Only 9% of population live in towns, 51,6 % of population live in alpine and mountain areas.

Demographic structure of population according to age (statistical data from 1996)		
age	Population in million	Percentage
0-14	8,67	41,3
15-24	4,03	19,2
25-34	2,9	13,8
35-44	2,14	10,2
45-54	1,55	7,4
55-64	1	4,8
65 and more	0,71	3,4
total	21	100,0

Resources

http://www.fncci.org
http://www.info-nepal.com/
http://www.info-nepal.com/dir/business/business.html
http://www.info-nepal.com/epb/
http://www.info-nepal.com/fips/
www.south-asia.com/news-ktmpost.html
www.un.org/Depts/tctd/escap/tctd/nepal.htm
www.south-asia.com/dotn/index.html
www.welcomenepal.com
www.mzv.cz/newdelhi

Authors:

Doc.RNDr. Bohdan Linda, CSc.
Doc.PaedDr. Jana Kubanová, CSc.
University of Pardubice
Faculty of Economics and administration
Department of mathematics
Studentská 95
53210 Pardubice
Czech Republic
e-mail: bohdan.linda@upce.cz
jana.kubanova@upce.cz

SROVNÁNÍ SYSTÉMŮ STATGRAPHICS PLUS, SAS A MS EXCEL PŘI VYHLEDÁVÁNÍ KVANTILŮ A HODNOT DISTRIBUČNÍCH FUNKCÍ

Tomáš Löster

Abstract:
The aim of the paper is to present some of advantages and disadvantages of systems STATGRAPHICS Plus, SAS, MS Excel. Attention is devoted to searching fractiles and distribution function values of different probability distribution in those systems and their confrotnation.

1. Úvod

Přechod z akademického roku 2004/2005 do akademického roku 2005/2006 na VŠE v Praze s sebou přináší změny ve výuce statistiky, mj. změnu softwarového produktu, který je využíván k výuce základních kurzů. Do letního semestru akademického roku 2004/2005 se při výuce používal systém STATGRAPHICS Plus a od zimního semestru následujícího akademického roku se postupně přechází na nový systém – SAS. Kromě těchto dvou produktů někteří vyučující využívají jako doplňkový systém MS Excel.

Cílem tohoto článku je poukázat na některé vlastnosti a ukázat odlišnosti ve vyhledávání kvantilů a hodnot distribučních funkcí těchto tří systémů.

2. Charakteristika systémů STATGRAPHICS Plus, SAS a MS Excel

Všechny tři produkty jsou vhodné pro použití jak v praxi, tak při výuce statistiky. Mají nabídkové uživatelské rozhraní, datový soubor je možné připravit přímo v rámci systému. Postupy sloužící k získání stejného výsledku se více či méně liší a každý systém má svá pozitiva a negativa.

STATGRAPHICS Plus – Mezi pozitiva tohoto produktu lze zařadit jednoduchou ovladatelnost, přehlednost, snadné pochopení výstupů, nápovědu, možnost práce s výstupy jako s obrázkem, který lze jednoduchými příkazy vložit do textového editoru, včetně grafů. Vstupní data je možné přenášet pomocí jednoduchých příkazů kopírování a vložení, formát desetinných čísel je kompatibilní s MS Excel, tj. desetinná místo je oddělené čárkou. Mezi nevýhody tohoto produktu patří zejména nestabilita při provádění složitějších metod – zejména u síťové instalace.

SAS – Mezi výhody tohoto produktu patří jeho snadná dostupnost na trhu, relativně značný výskyt v reálné praxi, formát výstupu provedených metod (HTML) atd. Mezi nevýhody tohoto produktu patří zejména složitější obsluha sloužící k získání jednoduchých výstupů, obtížná instalace produktu, velká náročnost z hlediska operačního systému (Windows XP) a velikosti místa na pevném disku.

MS Excel – je nejčastěji vyskytující se programový produkt, který lze využít k základním statistickým metodám. Výhodou tohoto produktu je téměř 100 % dostupnost. Obsluha tohoto systému je velmi jednoduchá a je vyučovaná již na středních školách. Mezi nevýhody tohoto produktu patří ne příliš vhodně zvolený překlad anglických termínů.

V některých případech se jedná o složeninu českého a anglického výrazu, např. VARVÝBĚR – jako jedna ze statistických funkcí.

3. Srovnání systémů z hlediska vyhledávání kvantilů a distribučních funkcí

STATGRAPHICS Plus – umožňuje velice jednoduchým způsobem vyhledat hodnoty kvantilů a hodnoty distribučních funkcí. Princip vyhledávání hodnot je u tohoto produktu založen na výběru z předem nadefinovaných menu nabídek. Před nalezením konkrétních hodnot kvantilů nebo hodnot distribučních funkcí si uživatel vybere z menu nabídky příslušný typ pravděpodobnostního rozdělení. Uživatel má na výběr reprezentanty z různých typů spojitých i nespojitých rozdělení. Mezi nespojitými rozděleními lze najít např. alternativní, binomické a Poissonovo rozdělení, nelze zde však nalézt rozdělení Hypergeometrické. Spojitá rozdělení zastupuje např. rozdělní normální, logaritmickonormální, exponenciální, studentovo, atd. Poté co si uživatel zvolí vhodný typ pravděpodobnostního rozdělení (zaškrtnutím vybraného typu a potvrzením), musí zadat parametry příslušného pravděpodobnostního rozdělení. U normálního rozdělení se jako druhý parametr zadává směrodatná odchylka. Po zadání vstupních hodnot (parametrů) si uživatel může pomocí menu nabídky vybrat, zda chce nalézt hodnoty kvantilů příslušného pravděpodobnostního rozdělení nebo hodnoty distribuční funkce.
Pro získání hodnoty kvantilu příslušného pravděpodobnostního rozdělení je nutné, aby uživatel pomocí menu nabídky zadal, o kolika procentní kvantil (x_p) se jedná. Tato hodnota je zadávána jako desetinné číslo, například ve formě 0,975 v případě, že uživatel hledá hodnotu 97,5 % kvantilu zvoleného pravděpodobnostního rozdělení. Uživatel má možnost získat až 5 hodnot kvantilů příslušného rozdělení najednou.
Při vyhledávání hodnot distribuční funkce, tedy dané pravděpodobnosti, musí uživatel kromě parametrů příslušného pravděpodobnostního rozdělení zadat také bod, ve kterém chce najít hodnotu distribuční funkce. Opět využije nabídku menu, ve které má možnost zadat až 5 bodů najednou, ve kterých chce hledat distribuční funkci. U nespojitých rozdělení je navíc nutné zohlednit definici distribuční funkce ve tvaru $F(x) = P(X \leq x)$. K získání výsledné hodnoty distribuční funkce je nutné sečíst dvě hodnoty, které nabízí výstup tohoto produktu. STATGRAPHICS Plus uvádí ve svých výstupech pouze ostré nerovnosti a samostatně rovnost (tedy hodnotu pravděpodobnostní funkce v příslušném bodě). Pro spojitá rozdělení tento problém vzhledem k jejich vlastnostem samozřejmě odpadá.

SAS – tento programový produkt, stejně jako STATGRAPHICS Plus umožňuje uživateli najít hledanou hodnotu distribuční funkce, případně hodnotu kvantilu příslušného pravděpodobnostního rozdělení. Uživatel si může vybírat z řady spojitých i nespojitých rozdělení, stejně jako v případě STATGRAPHICS Plus. Na rozdíl od STATGRAPHICS Plus vyhledávání těchto hodnot není založeno na výběru z menu nabídky, nýbrž uživatel zadává vstupní údaje pomocí kódu.
V případě, že uživatel hledá hodnotu kvantilu příslušného pravděpodobnostního rozdělení, musí v kódu využít následující syntaxi:

Kvantilová_funkce(P,<par_1,…,par_k>),

kde P je desetinné číslo, určující hodnotu 100*P% kvantilu a platí $P \in (0,1)$, par_1 – par_k jsou parametry příslušného pravděpodobnostního rozdělení.

Tabulka č.1: Příklady kvantilových funkcí

Pravděp. rozdělení	Kvantilová_funkce	par_1	par_2
Normální (normované)	PROBIT	p	-
Studentovo "t"	TINV	p	st.volnosti
Chí-kvadrát	CINV	p	st.volnosti

Pokud chce uživatel nalézt hodnotu např. 95% kvantilu normovaného normálního rozdělení, pak tedy musí užít následujícího kódu:

```
data _NULL_;
x=PROBIT(.95);
put x=;
run;
```

I v případě systému SAS, stejně jako u STATGRAPHICS Plus, je možné získat najednou několik hodnot kvantilů. Je však nutné postupně nadefinovat jednotlivá pravděpodobnostní rozdělení a % příslušných kvantilů a zadat příkaz pro určení těchto hodnot. Na rozdíl od systému STATGRAPHICS Plus je možné v jednom výstupu získat kvantily různých pravděpodobnostních rozdělení.

Vyhledávání hodnot distribučních funkcí je také založeno na zadávání údajů ve formě kódu. Podobně jako v případě vyhledávání kvantilů příslušného pravděpodobnostního rozdělení je při hledávání hodnot distribuční funkce třeba užít syntaxe, která má v tomto případě následující podobu:

CDF(‘rozdělení’,x, <par_1,…,par_k>),

kde ‘rozdělení’ představuje název příslušného typu pravděpodobnostního rozdělení používaný systémem SAS, x – bod, ve kterém uživatel hledá distribuční funkci a par_1 – par_k jsou parametry příslušného pravděpodobnostního rozdělení.

Tabulka č.2: Příklady pravděpodobnostních rozdělení pro vyhledávání distr. funkcí

Pravděp.rozdělení	'rozdělení'	par_1	par_2
Binomické	'BINOMIAL'	π	n
Normální (normované)	'NORMAL'	st.volnosti	-
Studentovo "t"	'T'	st.volnosti	-
Chí-kvadrát	'CHISQUARE'	st.volnosti	-

Pokud chce uživatel nalézt hodnotu distribuční funkce např. v bodě x = 1,645, předpokládá-li normované normální rozdělení, pak vstupní kód, který uživatel zadává má následující podobu:

```
data _NULL_;
x=('NORMAL',1.645);
put x=;
run;
```

I v tomto případě může uživatel získat najednou několik hodnot distribučních funkcí různých pravděpodobnostních rozdělení. Nejprve však musí nadefinovat typ

pravděpodobnostního rozdělení, dále body, v nichž hledá hodnoty distribučních funkcí a pak zadat příkaz k určení těchto hodnot.

MS Excel – umožňuje relativně jednoduchým způsobem vyhledávat hodnoty kvantilů a hodnoty distribučních funkcí různých pravděpodobnostních rozdělení. Uživatel si může vybrat z řady spojitých a nespojitých rozdělení. Mezi nespojitými lze nalézt například binomické, Poissonovo, hypergeometrické rozdělení, atd. Spojitá rozdělení jsou reprezentována například normálním, exponenciálním, logaritmickonormálním rozdělením atd. Na rozdíl od systémů STATGRAPHICS Plus a SAS umožňuje nalezení pouze jediné hodnoty v příslušné buňce.
Vyhledávání kvantilů je založeno na použití funkce, která má následující syntaxi:

*název*INV(par_1;…;par_k),

kde *název* představuje označení příslušného typu pravděpodobnostního rozdělení používané v MS Excel a par_1 – par_k jsou parametry definující příslušný typ pravděpodobnostního rozdělení a procento hledaného kvantilu.

Tabulka č.3: Příklady názvů a parametrů pro příslušná pravděp. rozdělení při hledání kvantilů

Pravděp.rozdělení	***název***	**par_1**	**par_2**	**par_3**	**par_4**
Normální (normované)	NORMS	P	–	-	-
Normální	NORM	P	stř.hodnota	sm.odch.	-
Studentovo "t"	T	Q	st.volnosti	-	-
Chí-kvadrát	CHI	R	st.volnosti	-	-

Kde: P - je desetinné číslo, určující hodnotu 100*P% kvantilu a platí $P \in (0,1)$.
Q - je hladina významnosti α, která se stanoví následujícím způsobem: při určení např. 95% kvantilu: $0,95 = 1 - \alpha/2 \Rightarrow \alpha = 0,1$
R - je hladina významnosti α, která se stanoví následujícím způsobem: při určení např. 95% kvantilu: $0,95 = 1 - \alpha \Rightarrow \alpha = 0,05$

Pokud chce uživatel nalézt hodnotu 95 % kvantilu normálního rozdělení s nulovou střední hodnotou a rozptylem rovno čtyřem, zapíše do určitého políčka tabulky výraz:

```
=NORMINV(0,95;0;2)
```

V případě požadavku uživatele najít několik hodnot kvantilů, je nutné vložit další funkci do nové buňky. Pro zjednodušení lze využít kopírování a změnit parametry resp. název.
Vyhledávání hodnot distribučních funkcí je v MS Excel, stejně jako hledání kvantilů, založeno na vytvoření funkce, která má v tomto případě následující syntaxi:

názevDIST(x;par_1;…;par_k),

kde *název* představuje označení příslušného typu pravděpodobnostního rozdělení používané v MS Excel, x – bod, ve kterém uživatel hledá distribuční funkci a par_1 – par_k jsou parametry pro příslušný typ pravděpodobnostního rozdělení.

Tabulka č.4: Příklady názvů a parametrů pro příslušná pravděp. rozdělení při hledání distr.funkcí

Pravděp.rozdělení	***název***	**par_1**	**par_2**	**par_3**
Normální (normované)	NORMS	-	–	-
Normální	NORM	stř.hodnota	sm.odch.	součet
Studentovo "t"	T	st.volnosti	strany	-
Chí-kvadrát	CHI	st.volnosti	-	-

Kde: součet – je zadáván jako 1 pro případ vyhledávání distribuční funkce
strany – tento parametr je zadáván podle toho, jak je vyjádřen příslušný kvantil. Pokud je vyjádřen jako $t_{1-\alpha}$ je hodnota parametru 1. Hodnota 2 se zadává v případě, že je kvantil vyjádřen jako $t_{1-\alpha/2}$.

Pokud uživatel chce zjistit, kolika procentní kvantil je hodnota 1,31 pro Studentovo rozdělení s 25 stupni volnosti, je nutné zadat funkci ve tvaru:

```
=TDIST(1,31;25;1)
```

Z výsledku je patrné, že se jedná o hodnotu 0,10, z čehož vyplývá, že se jedná o kvantil, jehož hodnotu určíme pomocí $1 - \alpha$, tedy $1 - 0{,}10 = 0{,}90$. Jedná se tedy o 90% kvantil.

4. Závěr

Z popisu jednotlivých systémů je patrné, že každý systém má své výhody i nevýhody. K vyhledávání kvantilů a hodnot distribučních funkcí se zdá nejjednodušší z hlediska ovládání systém STATGRAPHICS Plus a jako nejsložitější se jeví MS Excel. MS Excel ale na rozdíl od STATGRAPHICS Plus umožňuje získat hodnoty distribuční funkce hypergeometrického rozdělní (stejně jako systém SAS). Výhodou systému SAS je získání různých kvantilů a hodnot distribučních funkcí různých pravděpodobnostních rozdělení v jediném výstupu.

Literatura:

[1] ARLTOVÁ, M., BÍLKOVÁ, D., JAROŠOVÁ, E., POUROVÁ, Z.: *Příklady k předmětu statistika A,* VŠE, Praha 2003.
[2] JAROŠOVÁ, E., PECÁKOVÁ, I.: *Příklady k předmětu statistika B,* VŠE, Praha 2000.
[3] MAREK, L., a kol.: *Statistika pro ekonomy aplikace,* Profesional Publishing, Praha 2005.
[4] ŘEZANKOVÁ, H.: *Analýza kategoriálních dat,* VŠE, Praha 2005.

Kontaktní adresa:

Ing.Tomáš Löster
Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky
nám. W. Churchilla 4
130 67 Praha 3
Česká republika
losterto@vse.cz

Reprezentatívnosť vo výskumoch verejnej mienky

Ján Luha

1. Úvod

Všeobecne možno definovať pojem reprezentatívnosti ako zhodu distribučných funkcií skúmaných výberových charateristík s odpovedajúcimi im distribúciami v záujmovom základnom súbore. V závislosti od skúmaných znakov (charakteristík) a od spôsobu vytvorenia výberového súboru možno pojem reprezentatívnosti ďalej bližšie špecifikovať. Budeme sa zaoberať kvalitatívnymi znakmi a dvoma spôsobmi konštrukcie výberových súborov: 1. náhodný výber a 2. kvótový výber. Potom definujeme reprezentatívnosť ako:

1. nevychýlenosť výberových odhadov

2. zhodu vybraných parametrov znakov základného a výberového súboru (napr. vybrané demografické znaky ako pohlavie, vek, vzdelanie, ...).

Medzi podstatné faktory ovplyvňujúvce reprezentatívnosť patria: vhodnosť opory výberu, pokrytie, použitá výberová metóda a vo veľkej miere návratnosť. Reprezentatívnosť (do značnej miery) nesúvisí s rozsahom výberu. Rozsah výberu ovplyvňuje najmä presnosť výsledkov.

Vo všeobecnosti sa predpokladá, že vysoká návratnosť zaručuje aj reprezentatívnosť, pritom sa kladie malá pozornosť skladbe návratnosti. Aj pri vysokej návratnosti môže prísť ku porušeniu reprezentatívnosti, ak získaný výberový súbor neodpovedá štruktúre základného súboru. Aby sme dokázali overiť zhodu štruktúry parametrov výberového a základného súboru, musíme mať určené relevantné znaky (vzhľadom na skúmanú problematiku), inak nedokážeme posúdiť, či výberový súbor je reprezentatívny. Žiaľ, častokrát sú takého charakteristiky neznáme.

Obmedzíme sa na všeobecné východiská a preto sa nebudeme zaoberať konkrétnymi výberovými procedúrami ako napr. oblasné výbery a pod.

2. Reprezentatívnosť pri náhodnom výbere

Ako je uvedené vyššie, reprezentatívnosť v tomto prípade definujeme ako nevychýlenosť, resp. nestrannosť výberového odhadu:
Nech X je skúmaný znak a μ je jeho stredná hodnota.
Ak X_1, X_2, ... X_n je realizácia náhodného výberu pre skúmaný znak X, tak ako odhadovú štatistiku pre strednú hodnotu uvažujeme výberový priemer

$$X' = \sum_i Xi / n$$

potom nevychýlenosť pre odhad zvolenej štatistiky – priemeru znamená, že stredná hodnota

$$E(X') = \mu.$$

Mierou vychýlenia je odchýlka $\theta = X´ - \mu$.

Jednou zo základných vlastností náhodného výberu je nevychýlenosť a preto **v prípade realizácie náhodného výberu je táto realizácia reprezentatívna** v zmysle prvej definície. Získaný výberový súbor však musí byť náhodný a teda nevychýlený. Pri takmer 100% návratnosti môžeme predpokladať, že výberový súbor je reprezentatívny**.**
Na overenie reprezentatívnosti, v prípade náhodného výberu, keď nepoznáme rozdelenie skúmaných znakov, alebo vhodné charakteristiky rozdelenia potrebné na testovanie, môžeme využiť druhú definíciu reprezentatívnosti. Teda taktiež za predpokladu, že máme k dispozícii rozhodujúce relevantné znaky a poznáme ich rozdelenie v základnom súbore.

3. Reprezentatívnosť pri kvótovom výbere

Ako je už uvedené v časti 1, v tomto prípade je reprezentatívnosť definovaná ako zhoda parametrov znakov základného a výberového súboru.

Uvažujme napríklad znaky pohlavie a vzdelanie. Nech rozdelenie relatívnych početností uvažovaných znakov v základnom a výberovom súbore je ako ukazuje nasledujúca tabuľka. Ak test zhody (napr CHí – kvadrát) nezistí signifikantné rozdiely, tak z hľadiska týchto znakov je výberový súbor reprezentatívny.

Reprezentatívnosť výberového súboru				
Znak	Základný súbor	Výberový súbor	Rozdiel %	CHI2 Test, HV 5 %
Pohlavie:				rozdiely nie sú
muž	47,91	47,78	-0,13	štatisticky
žena	52,09	52,22	+0,13	významné
Vzdelanie:				rozdiely nie sú
základné	26,81	26,60	-0,21	štatisticky
stredné bez maturity	30,22	30,60	+0,38	významné
stredné s maturitou	32,88	32,55	-0,33	
vysokoškolské	10,09	10,25	+0,16	

Na overenie reprezentatívnosti v tomto prípade možno tiež použiť Fisherov exaktný test, Waldovu aproximáciu, ktoá je spomínaná v nasledovnej časti a pre overenie zhody podielov aj z-test.

4. Rozsah výberového súboru

Rozsah výberovej vzorky nie je síce determinujúci pre reprezentatívnosť, ale je zase podstatný pre spoľahlivosť výsledkov výberového zisťovania. Pripomeňme si, že pri jednoduchom náhodnom výbere a pre kvalitatívne znaky ho stanovujeme podľa vzorca:

$$\mathbf{n=(1,96/d)^2 p.q}$$

pri úrovni významnosti 5% (za predpokladu npq > 9) a keď uvažujeme znak s odpoveďami (áno / nie), kde q=1-p a d je predpokladaná presnosť odhadu podielu p výskytu hodnoty áno.

Ak uvažujeme výber bez opakovania zo základného súboru o veľkosti N, tak vzorec na výpočet rozsahu výberu má tvar:

$$n=(1,96/d)^2 pq\ (1 - f),$$

kde f=n/N je výberový pomer.
Pritom však musíme sledovať splnenie podmienky aplikácie uvedeného vzorca a síce npq > 9. Pre orientáciu uvádzame tabuľku minimálnych rozsahov vyplývajúcu z uvedeného vzťahu pre rôzne hodnoty podielu (Poznámka: Pre p a 1-p sú tieto hodnoty zhodné):

Tabuľka č. 1: Minimálne zozsahy výberu pre normálnu aproximáciu

p	0,01	0,02	0,05	0,10	0,15	0,20	0,30	0,50
p	0,99	0,98	0,95	0,90	0,85	0,80	0,70	0,50
n	**909**	**459**	**189**	**100**	**71**	**56**	**43**	**36**

Pre rýchlu aproximáciu spoľahlivosti výsledkov výskumu používame často odhad pre najširší interval spoľahlivosti. Ten dosiahneme v prípade p=0.5. Vtedy platí približne

$$d=98/\sqrt{(n)}.$$

Pre rozsahy výberov obvykle používané vo výskumoch verejnej mienky je teda možné použiť na meranie spoľahlivosti hore uvedený vzorec. Pre rýchlu orientáciu možno zostaviť tabuľku :

Tabuľka č.2: Odhad reliability pre vybrané rozsahy výberu

n	400	500	600	700	800	900	1000	1100	1200	1300	1400
d (v %)	4,9	4,4	4,0	3,7	3,5	3,3	3,1	3,0	2,8	2,7	2,6

Vzhľadom na zaužívanú konvenciu prezentujeme spoľahlivosť d v tejto tabuľke tiež v percentách. Pri rozsahu výberovej vzorky nad 1000 osôb je výberová chyba menšia než 3%.

Na spresnenie by bolo, najmä pri malých rozsahoch výberu, potrebné použiť exaktné vzorce. Pre praktické účely možno použiť (Agresti and Coull (1998)) odporúčajú modifikovanú Waldovu metódu. Podľa nej sa počíta modifikovaný podiel a modifikovaná spoľahlivosť:

$$p´ = (x + 2)/(n + 4) \text{ a } d´ = 1,96.\sqrt{((p´(1-p´)/(n + 4))},$$

Hoci sa vedú spory o možnosti využiť vzoce platné pre náhodný výber aj pre kvótový výber v praxi sa tieto vzorce využívajú. ***Existujú publikácie, ktoré ilustrujú možnosť aplikácie uvedených vzorcov.*** Viď. napr. Anděl M., Černý R., Charamza P., Neustadt J. (2004), Gourieroux Christian (1987).

5. Porušenie reprezentatívnosti

V prípade porušenia reprezentatívnosti, teda nevychýlenosti výberového odhadu v prvom koncepte reprezentatívnosti, resp. signifikantnom rozdiele pri niektorom kontrolovanom demografickom znaku, je potrebné v závislosti od veľkosti porušenia reprezentatívnosti, urobiť opatrenia na objektivizáciu získaných výsledkov.

Všeobecne je to komplexný problém a nie je možné ho jednoducho opísať. Spomenieme aspoň základné zdroje chýb pri realizácii výberovej procedúry:
A: „Nezhoda“ medzi základným súborom (populáciou) a oporou výberu.
Pre realizáciu náhodného výberu je podstané aby bola k dispozícii kvalitná opora výberu – teda zoznam, najlepšie elektronický, prvkov základného súboru. V prípade kvótového výberu potrebujeme aktuálne údaje o štruktúre kontrolovaných demografických znakov v základnom súbore.
B: „Nezhoda“ medzi oporou výberu a výberovým súborom (teoreticky definovaným).
Táto výberová chyba indukuje aj chybu pre získaný výberový súbor
C: „Nezhoda“ medzi výberovým súborom a získaným výberovým súborom.
Táto chyba je spôsobená menšou návratnosťou a aj skladbou návratnosti.

V prípade náhodného výberu je obvyklé realizovať doplnkové zisťovanie, prípadne resamling, alebo váženie, ak môžeme zvoliť „najdôležitejšie“ znaky, podľa ktorých bude možné realizovať odpovedajúce procedúry. Pri kvótovom výbere je – v závislosti veľkosti porušenia reprezentatívvnosti – možné objektivizovať odhady pomocou váženia, resamplingu, alebo aj doplnkového zisťovania.

Najpoužívanejším testom zhody je X^2 test. Okrem testovacích štatistík založených na veličine Chí-kvadrát možno použiť Fisherov exaktný test. Zo známych štatistických softvérov ho obsahuje modul Exact tests SPSS. No overenie zhody podielu znaku vo výberovom a základnom súbore možno použiť aj z-test (ak sú splneneé podmienky aproximácie, ako sú uvedené v časti 4 a Tabuľke 1.

kde:
Ps = podiel vo výberovom súbore,
P = podiel v základnom súbore,
n = rozsah výberu.

Táto štatistika má asymtoticky N(0,1) zozdelenie.

6. Literatúra

1. Agresti A., Coull Ba.: Approximate is better than "Exact" for interval estimation of binomial proportions. The American Statistician. 52:119-126, 1998.
2. Anděl M., Černý R., Charamza P., Neustadt J.: Přehled metod odhadu statistické chyby ve výběrových šetřeních. Informační bulletin České statistické společnosti. Číslo2-3/2004.
3. Cochran G.W.: Sampling Techniques. Third Edition. Wiles, New York 1977.
4. Čermák V.: Výběrové statistické zjišťování. SNTL/ALFA Praha 1980.
5. Elliott, C. and Ellingworth, D. (1997) 'Assessing the Representativeness of the 1992 British Crime Survey: The Impact of Sampling Error and Response Biases' *Sociological Research Online*, vol. 2, no. 4, http://www.socresonline.org.uk/socresonline/2/4/3.html
6. Gourieroux Christian: Théorie des Sondages. INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES, Economica. 1987.
7. Herzmann J., Novák I., Pecáková I.: Výzkumy veřejného mínění. Skriptum VŠE Praha, Fakulta informatiky a statistiky, 1995.
8. Chajdiak J.: Štatistika jednoducho. Statis Bratislava 2003.
9. Kanderová, M. – Úradníček, V.: Aplikovaná štatistika vo finančno-ekonomickej praxi.

OZ Financ, Banská Bystrica 2005.
10. Kanderová, M. – Úradníček, V.: Štatistika a pravdepodobnosť pre ekonómov. 1. časť. OZ Financ, Banská Bystrica 2005.
11. Kaščáková, A. – Nedelová, G. Možnosti popisu štatistického súboru podľa typu sledovaného znaku. In. Výpočtová štatistika. Zborník príspevkov z 12. medzinárodného seminára. Bratislava: Slovenská štatistická a demografická spoločnosť, 2003, s. 46-49.
12. Kol.: Názory občanov Slovenska na životné prostredie. Štatistické analýzy a informácie. ŠÚ SR, júl 1995.
13. Likeš J., Laga J.: Základní statistické tabulky. SNTL, Praha 1978.
14. Luha J.: Testovanie štatistických hypotéz pri analýze súborov charakterizovaných kvalitatívnymi znakmi. STV Bratislava, 1985.
15. Luha J.: Meranie spoľahlivosti výsledkov výskumu verejnej mienky. Zborník príspevkov Využitie štatistických metód v sociálno - ekonomickej praxi, EKOMSTAT´94 29.5.-3.6. 1994 Trenčianske Teplice, SŠDS.
16. Luha J.: Exaktné intervaly spoľahlivosti pre podiely. Slovenská štatistika a demografia 1/96.
17. Luha J.: Matematickoštatistické aspekty spracovania dotazníkových výskumov. Štatistické metódy vo vedecko-výskumnej práci 2003, SŠDS, Bratislava 2003.
18. Mace A. E.: Samle-size determination. Reihold. New York 1964.
19. Motulsky H.: GraphPad PRISM, Version 4.0 Statistical Guide. Statisstical analyses for laboratory and clinical research. 2005. GraphPad Software, Inc. http://www.graphpad.com/
20. Pecáková I.: Statistické aspekty terénních pruskumu I. Skriptum VŠE Praha 1995.
21. Príručky SPSS. Ver. 13.

RNDr. Ján Luha, CSc.
Ústav pre výskum verejnej mienky pri SÚ SR
(Adresa sídla: Hanulova 5/c, Bratislava)
Poštová adresa: Miletičova 3, 824 67 Bratislava
Tel: 02/69250 242
E-mail: Jan.Luha@statistics.sk

Softvérové riešenie úloh zhlukovej analýzy

Silvia Megyesiová

Abstract: The article deals with comparison of statistical software using by cluster analysis. In the article we compare three statistical software – SAS-Enterprise Guide, SPSS for Windows 13.0 and NCSS2001 Test. According to our comparison we recommend to use for cluster analysis software SPSS and NCSS, because the system SAS-EG has some disadvantages compared to other two statistical software.

Kľúčové slová: štatistický softvér, zhluková analýza

Použitie štatistického softvéru pri riešení úloh viacrozmernej analýzy sa stáva v súčasnosti samozrejmosťou. Už počas štúdia sa študenti vysokých škôl oboznamujú s prostredím štatistických softvérov ako i s ich používaním. Zhluková analýza je obľúbená metóda rozkladu súboru pozorovaných štatistických jednotiek na niekoľko relatívne homogénnych podsúborov, zhlukov a to tým spôsobom, aby si jednotky (objekty) v rámci konkrétneho zhluku boli čo najviac podobné a naopak štatistické jednotky, patriace do rôznych zhlukov si boli podobné čo najmenej.

Každá štatistická jednotka je popísaná skupinou štatistických znakov (premenných, ukazovateľov). Úlohou zhlukovej analýzy je teda zaradenie jednotiek do skupín, tried, pričom nepotrebujeme poznať predpoklady a informácie o tvare a type viacrozmerného rozdelenia.

Vytvorené zhluky by mali byť kompaktné a navzájom relatívne izolované, teda jednotky z toho istého zhluku by si mali byť podobné, kým jednotky pochádzajúce z rôznych zhlukov by sa svojimi vlastnosťami mali čo najviac líšiť.

V súčasnosti je na pôde Ekonomickej univerzity v Bratislave používaná univerzitná verzia štatistického softvéru SAS: SAS-Enteprise Guide 2.1. Tento softvér v porovnaní so štatistickými softvérmi SPSS ako i NCSS ponúka najnižší komfort z pohľadu tvorby modelov zhlukovej analýzy.

Profesionálne štatistické softvérové produkty – SPSS 13.0 for Windows ako i NCSS 2001 Test - síce nie sú oficiálne dostupné na pracovisku Ekonomickej univerzity v Bratislave, avšak tieto produkty je možné nainštalovať z internetových stránok spoločností, ktoré ich ponúkajú na bezplatné otestovanie užívateľmi na časovo obmedzené obdobie.

Používaním troch vyššie spomenutých softvérových produktov by sme chceli poukázať na ich silné ako i slabé stránky jednotlivých produktov, ku ktorým sme dospeli pri tvorbe modelov zhlukovej analýzy.

- ***SAS – Enterprise Guide***

Konkrétne výhrady použitia SAS-EG pri tvorbe modelov zhlukovej analýzy sú nasledovné:

- dialogové okno tvorby modelu neumožňuje štandardizáciu pôvodne zistených údajov, to znamená, že štandardizáciu ukazovateľov zistených v rôznych merných jednotkách nie je možné uskutočniť priamo v procese tvorby modelu zhlukovej analýzy, ale je potrebné túto štandardizáciu uskutočniť pomocou príkazov v inom dialógovom okne,
- v dialógovom okne tvorby modelu nie je možné voliť miery vzdialeností objektov, prednastavená je štvorcová Euklidovská vzdialenosť, v prípade potreby zmeny metriky je nutné zistiť možnosť zmeny vygenerovaného programu a zápisu konkrétneho príkazu na výpočet zvolenej metriky vzdialeností, tento nedostatok možnosti priamej zmeny metriky v dialógovom okne tvorby modelu zhlukovej analýzy sa nám javí ako výrazné skomplikovanie použitia danej procedúry vzhľadom k tomu, že súčasný používatelia Windows verzií sotvérových produktov neovládajú programovacie postupy a naopak očakávajú istý komfort pri používaných takýchto produktov,
- výber zhlukovacích procedúr v danej verzii SAS-EG taktiež nepovažujeme za postačujúce, k dispozícii je výber štyroch procedúr, z toho troch hierarchických a jednej nehierarchickej zhlukovacej procedúry:
    - metóda priemernej väzby
    - centroidná metóda
    - Wardova metóda
    - metóda k-priemerov

Výhody používania SAS-EG k tvorbe modelu zhlukovej analýzy:

- ako hlavnú výhodu použitia tohto produktu vidíme v tom, že je oficiálne dostupný tak zamestnancom ako i študentom EU v Bratislave ,
- jednoduchá možnosť importovania údajov, ktoré boli pôvodne vytvorené v inom prostredí, ako prípad môžeme uviesť jednoduchý spôsob importovania údajov vytvorených v tabuľkovom procesore Microsoft Excel,

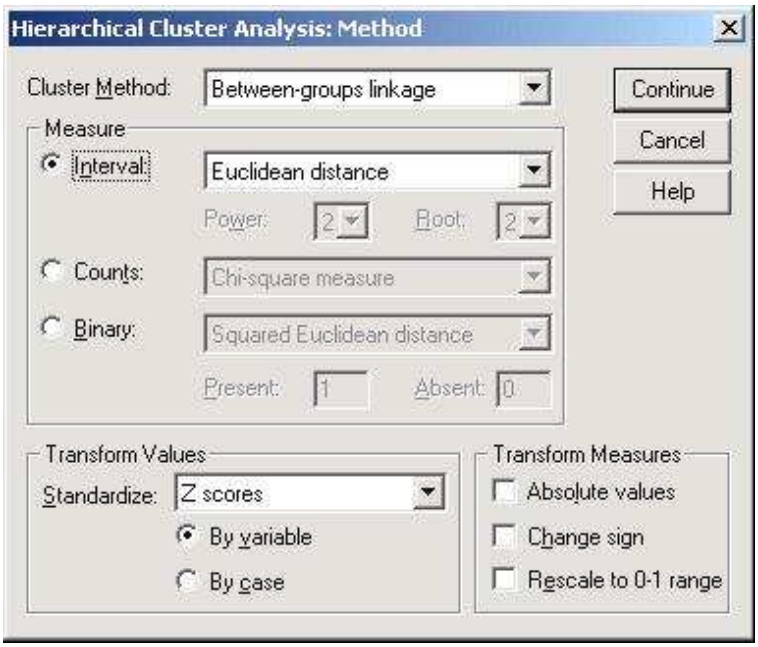
- výpočet koeficientov determinácie a semiparciálnych koeficientov determinácie na jednotlivých stupňoch zhlukovania.

➢ ***SPSS 13.0 for Windows***

Použitie vyššie uvedeného štatistického softvéru k tvorbe modelov zhlukovej analýzy má v porovnaní so SAS-EG niekoľko výhod:

- v prvom rade je tvorba modelov zhlukovej analýzy rozdelená do dvoch procedúr, jedna z týchto procedúr *(Analyze→Classify→Hierarchical Cluster)* umožní vytvoriť hierarchické modely zhlukovej analýzy, kým druhá procedúra *(Analyze→Classify→K-Means Cluster)* je určená na tvorbu nehierarchických modelov zhlukovej analýzy.
- výhodou použitia tohto softvéru k tvorbe modelov zhlukovej analýzy je i v tom, že v prípade potreby je možné priamo v tomto dialógovom okne uskutočniť štandardizáciu pôvodne zistených ukazovateľov, ktoré sú často namerané v rôznych merných jednotkách. Táto trasformácia sa uskutoční pomocou voľby požadovanej transformácie na spodnej časti dialógového okna *(Transform Values)* – viď obr. 1,

Obrázok 1 Dialógové okno tvorby hierarchických modelov zhlukovej analýzy v SPSS 13.0 for Windows



- ďalšou výhodou tohto softvéru oproti SAS-EG je jednoduchá voľba metriky vzdialeností objektov *(Measure)*, ktorá sa v prípade použitia SAS-EG sa mohla meniť

iba pomocou preprogramovania vygenerovaného kódu tvorby zhlukovej analýzy, čo pokladáme z hľadiska komfortu analýzy za veľmi nepraktické. Vybrať si môžeme zo širokej ponuky mier vzdialeností, nepodobností objektov (Euklidovská vzdialenosť, štvorcová euklidovská vzdialenosť, Pearsonov koeficient korelácie, Chebychevova vzdialenosť, Manhattan-Block, Minkovského vzdialenosť) ako i z mier podobnosti objektov, ktoré sú určené k porovnávaniu objektov nemetrického charakteru (binárnych premenných),

- konkrétnou voľbou v *Cluster Method* si môžeme vytvoriť celkovo 7 hierarchických modelov zhlukovej analýzy:
    - metóda priemernej väzby – between-groups linkage, UPGMA,
    - metóda priemernej väzby – within-group linkage, WPGMA,
    - metóda najbližšieho suseda,
    - metóda najvzdialenejšieho suseda,
    - centroidná metóda,
    - mediánová metóda,
    - Wardova metóda

  (v prípade SAS-EG bolo možné vybrať si iba z troch ponúknutých hierarchických procedúr).

  Túto širokú ponuku rôznych hierarchických procedúr v SPSS for Winfows taktiež považujeme za výhodu daného produktu oproti softvéru SAS-EG.

Funkčnú verziu tohto softvéru SPSS for Windows 13.0 je možné nainštalovať na osobný počítač na časovo obmedzené obdobie priamo z internetovej stránky spoločnosti SPSS Inc.: *www.spss.com*.

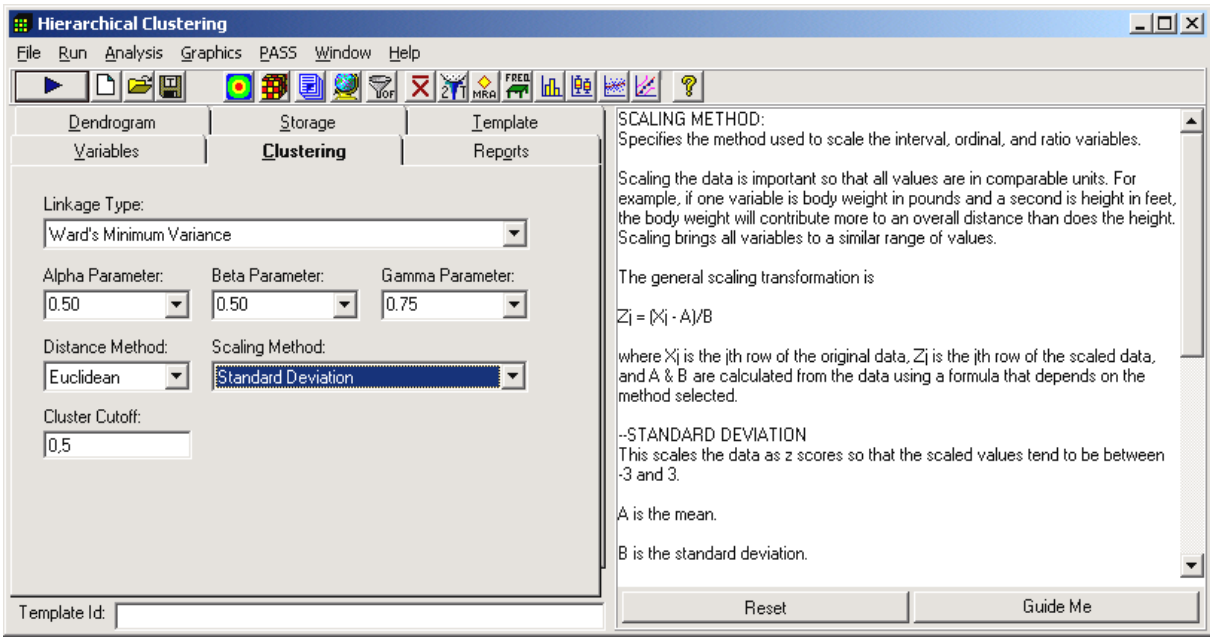
- ***NCSS2001 Test***

Tento štatistický softvér je obľúbený pre svoju jednoduchú manipuláciu ako i širokú škálu ponúkaných analýz. Jeho výhodou je i pomoc *(Help)* objavujúca sa v pravej časti dialógového okna. Táto pomoc sa objaví po vyvolaní ľubovolného dialógového okna a týka sa teda iba nápovedy k danej problematike, nápovedy k riešeniu konkrétnej štatistickej analýzy.

Dialógové okno tvorby hierarchických modelov zhlukovej analýzy je znázornené na obr. 2. V pravej časti tohto okna vidíme nápovedu, zobrazenú pre štandardizáciu premenných *(Scaling Method)*, v tomto okne môžeme pomocou jeho rolovania získať bližšie informácie o možnostiach štandardizácie premenných, o vzťahoch používaných pri jednotlivých typoch

štandardizácie. Táto nápoveda sa nám zdá byť pre užívateľa veľmi vhodným prostriedkom pri práci s daným štatistickým softvérom.

Obrázok 2 Dialógové okno tvorby modelov zhlukovej analýzy v NCSS



Nápovedu je možné zobraziť jednoduchým pohybom kurzoru myši v priestore dialógového okna, nápoveda sa mení podľa toho, kde kurzor myši umiestnime.

Výhodou používania NCSS 2001 k tvorbe modelov zhlukovej analýzy:

- tvorba modelov je rozdelená do niekoľkých procedúr:
  - hierarchické modely zostavuje procedúra: *Analysis→Clustering→Hierarchical*,
  - nehierarchické modely tvorí procedúra: *Analysis→Clustering→K-Means*,
  - procedúry založené na medoidoch, optimálnych stredoch zhlukov sa tvoria procedúrou: *Analysis→Clustering→Medoid Partitioning*,
  - Fuzzy zhluková analýza, ktorá umožňuje zhlukovanie jedného objektu do viac než jedného zhluku sa stanoví procedúrou: *Analysis→Clustering→Fuzzy*.

  Tento štatistický softvér teda s porovnaním s SPSS ako i s SAS-EG poskytuje najširšiu možnosť výberu rôznych procedúr, ktoré tvoria modely zhlukovej analýzy, ide hlavne o rozšírenie modelov o fuzzy zhlukovacie procedúry ako i procedúry založené na optimálnych stredoch, teda medoidoch. Obe tieto procedúry neboli súčasťou žiadneho iného štatistického softvéru použitého v tejto práci,

- štandardizácia premenných sa uskutočňuje priamo v dialógovom okne tvorby modelu zhlukovej analýzy, podobne ako v prípade SPSS,
- poskytnutie rýchlej a jednoduchej nápovedy v procese tvorby modelov, táto rýchlo dostupná pomoc pri zhlukovaní nie je dostupná v softvéri SPSS ani v SAS-EG,
- práca s formátovaním teda úpravou vytvorených grafov, dendogramov je veľmi jednoduchá, grafy a tabuľky sa dajú taktiež jednoducho kopírovať napríklad do MS-Wordu.

Záverom hodnotenia práce s tromi vyššie uvedenými štatistickými softvérmi, pomocou ktorých je možné tvoriť modely zhlukovej analýzy musíme konštatovať obmedzené možnosti použitia softvéru SAS-EG a ako vhodnejší štatistický softvér v danom prípade musíme odporučiť softvér SPSS ako i NCSS.

Príspevok je súčasťou riešenia úlohy VEGA 1/2555/05.

Literatúra:

[1] AJVAZJAN, S. - BEŽAJEVOVÁ, Z. - STAROVEROV, O.: Metody vícerozměrné analýzy, Praha, SNTL, 1981

[2] HEBÁK, P. - HUSTOPECKÝ, J.: Vícerozměrné statistické metody s aplikacemi, Praha, SNTL, 1987

[3] KŘIVÁNEK, M.: Algirithmic and Geometric Aspects of Cluster Analysis, Praha, Academia, 1991

[4] LUKASOVÁ, A. – ŠARMANOVÁ, J.: Metody shlukové analýzy, Praha, SNTL, 1985

[5] MEGYESIOVÁ, S.: Porovnanie krajín Európskej únie pomocou viackriteriálneho hodnotenia variant, Forum Metricum Slovacum, TOM VIII., s. 79-86, SŠDS, Bratislava 2004

[6] www.ncss.com

[7] www.spss.com

Ing. Silvia Megyesiová
Ekonomická univerzita v Bratislave
Podnikovohospodárska fakulta Košice
Tajovského 13
040 00 Košice

megyesiova@euke.sk

Porovnanie vývoja HTFK vo vzťahu ku Hrubému domácemu produktu
The comparison of GFCF to relation of Gross Domestic Produc

Andrej Mikuš

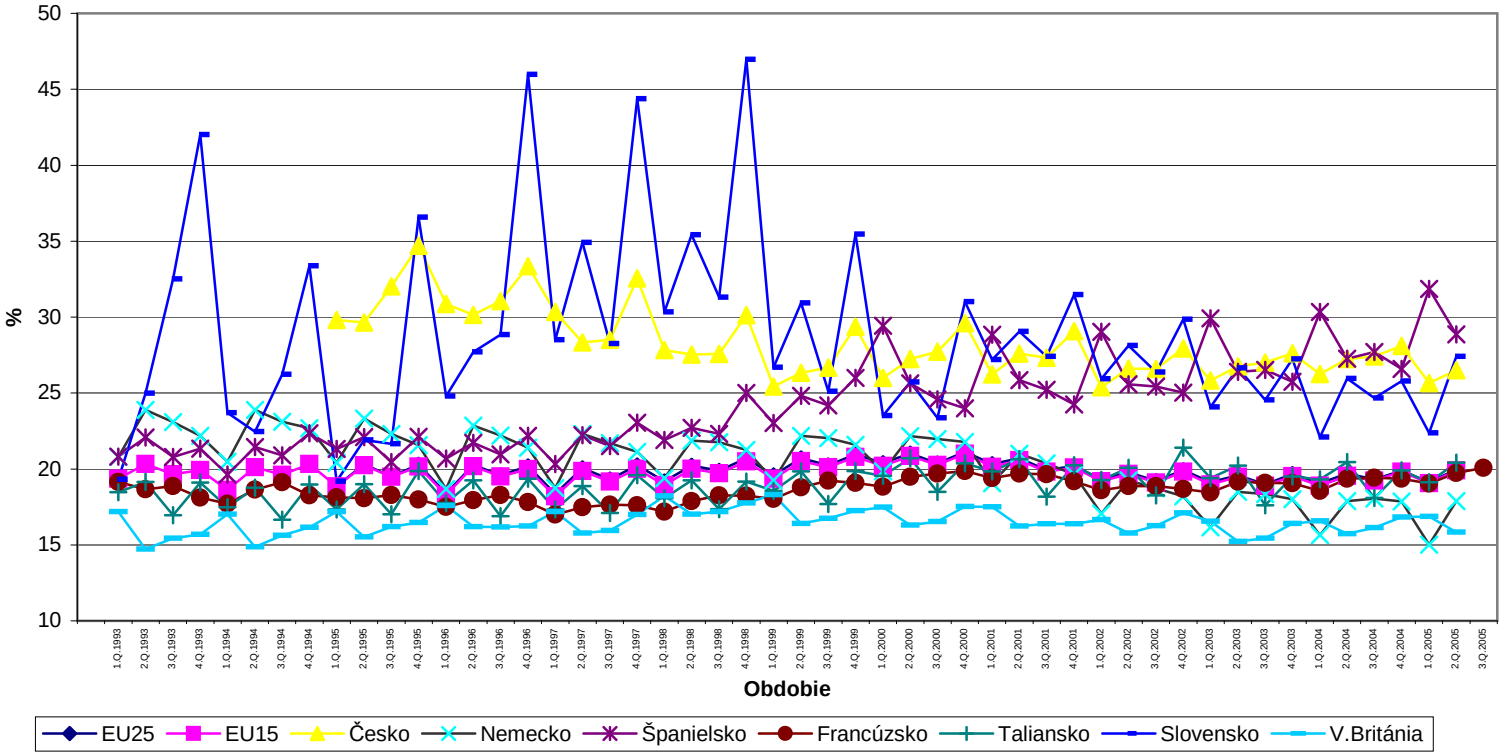
Zdroje údajov:Eurostat
Andrej Mikuš, Ing., Andrej. Mikus@statistics.sk

The influence of GCFC on whole Domestic product, their components like machinery equipment and its role in Gross Fixed Capital Formation.

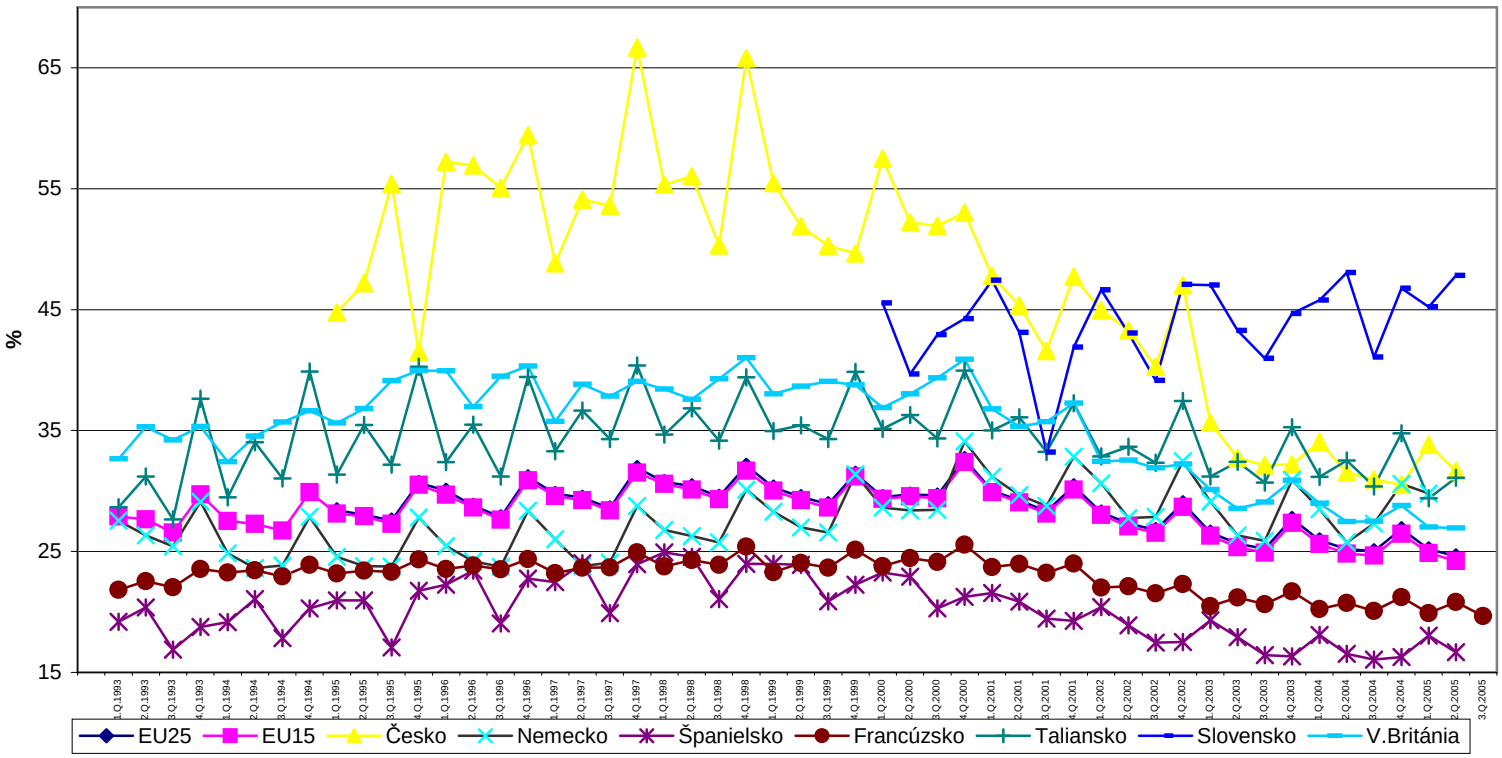
Z hľadiska objemov a ďalšieho vývoja Hrubého domáceho produktu (HDP) má významnú úlohu a postavenie Hrubá tvorba fixného kapitálu (HTFK). Ide o prostriedky vynaložené na obstaranie nových investícií včítane odpisov. Preto z pohľadu ďalšieho HDP je potrebné skúmať podiel kapitálu ku HDP a takisto je zaujímavé skúmať štruktúru kapitálu podľa jednotlivých produktov. Tu sa jedná o to či sa kapitálové investície nadobúdajú vo forme pôdohospodárskych investícií, kovových výrobkov, strojov a zariadenia, výstavby bytov a domov, ostatných budov a stavieb a či do dopravných prostriedkov.

Na porovnanie som vybral: štáty za celú EÚ –25; pôvodné členské štáty EÚ-15; Česko; Nemecko; Španielsko; Francúzsko; Taliansko; Slovensko a Veľkú Britániu. Na porovnanie som použil objemy, ktoré sú vyjadrené v bežných cenách v Euro.
Ako prvým skúmaným faktorom bol podiel Hrubej tvorby fixného kapitálu na HDP. Ako vidieť z grafu č.1. Najvýraznejší rast podielu HTFK (v porovnaní ku HDP) bol za EÚ-25 zaznamenaný v období 1998-2000, pričom najväčší podiel na celkovom prírastku malo Francúzsko, kde bol rast od 3 do 4%, Španielskom medzi 7 až 8% ako aj Taliansko, ktoré dosahovalo 4 až 5% ročné prírastky podielov pričom je vidieť aj extrém u Španielska v 1.Q.2000. Hlavnou brzdou kapitálových investícií sa z uvedeného javí vývoj v Nemecku, kde bol aj podiel na HDP malý, aj sa za celé obdobie skôr zaznamenával pokles a to hlavne v rokoch 2001 až 2003. V Česku tiež klesal podiel ako aj dochádzalo k poklesu podielu HTFK na celkovom HDP.
Možno totiž povedať, čím je sklon ku spotrebe väčší, tým menej sa z celkového HDP investuje a viac ide na konečnú spotrebu domácností, alebo sa prejavuje v zápornom salde zahraničného obchodu. Slovensko z HTFK bolo na výslní v rokoch 1996 až 1998, kde bol dosahovaný rast 33%, pričom v období rokov 1999, 2000, 2002 a 2003 s výnimkou roka 2001 bol zaznamenaný hlboký prepad celkového podielu na HDP. Medziročný pokles v niektorých kvartáloch roku 1999 bol až 20-25%. Čo sa týka podielu, tak Slovensko dosahovalo v uvedených rokoch od 25 do 35%, pričom v niektorých kvartáloch až vyše 45%, čo dobre vidieť z tab.č.1 ako aj z grafu č. 1. Vyspelé štáty dosahujú spravidla hodnoty oscilujúce okolo 20%, pričom Veľká Británia dokonca hodnoty len o niečo viac 15 %.
Druhým pohľadom je podiel kovových výrobkov a zariadenia na celkových investíciách. Z nízkeho podielu kovových výrobkov, strojov a zariadenia možno usudzovať o ekonomike, ktorá viac investuje do výstavby domov a bytov, dopravných prostriedkov – teda do sféry služieb ako do primárnej produkcie v priemysle. Tým však podporuje väčší sklon ku spotrebe, a tým ťahá celkové investície nadol. Z grafu č.2. vidieť, že podiel za EU - 25 ako aj EU – 15 osciluje o 30%, pričom došlo k miernemu poklesu v priebehu skúmaného obdobia. Možno povedať, že10 nových členských štátov na celkový objem investícií má len nepatrný vplyv. Určitým pozitívnym prvkom je rast podielu strojov a zariadenia, z celkových investícií, v Nemecku z oblasti 25% do oblasti 30%. Najväčší pokles zaznamenalo Česko, ktoré pokleslo z podielu v niektorých kvartáloch 65% na hranicu 30%. Na Slovensku sme začali sledovať uvedenú štruktúru rozdelenia HTFK od roku 2000, takže časová rada nepredstavuje relatívne krátky úsek, ale daný podiel osciluje okolo 45%. V Španielsku sa tento podiel nachádzal v rozmedzí 20% v roku 1993, potom sa postupne tento podiel zvyšoval až na 24% v rokoch 1997 a 1998 a potom sa pozvoľna vrátil na pôvodné hodnoty 18-19%, pričom počnúc rokom 2003 sa hodnoty priblížili hodnotám 16%. Vo Francúzsku bol zaznamenaný nárast z 21-23% v roku 1993 s vrcholom v rokoch 1998 až 2001, pričom v roku 2004 tieto hodnoty už boli na 20% a v tomto roku dokonca klesli na 19%.V Taliansku sa vývoj príliš nemenil počas celého obdobia a dosahoval hodnoty okolo30-35%. Výnimkou boli roky 1997-2000, kde sa pohyboval podiel na hranici 40%.

Graf č. 1. Porovnanie podielu HTFK na HDP v jednotlivých členských štátoch EU



Graf č. 2. Podiel strojov a zariadenia na celkovom HTFK



Testy nezávislosti v řídkých kontingenčních tabulkách

Iva Pecáková

Abstract: By testing independence in two-way contingency tables, usually the Pearson statistic or the likelihood ratio chi-squared statistic is used. When cell counts are too small to permit chi-squared approximations, it is required to use an alternative approximation for the distribution of the statistic, or an exact test.

Výběrové četnosti n_{ij} v dvourozměrné kontingenční tabulce (tabulka 1) lze chápat jako nespojité náhodné veličiny a k provádění induktivních úsudků o populaci využít vlastností výběrových rozdělení četností či jejich funkcí.

Tabulka 1: Dvourozměrná kontingenční tabulka

	b_1	b_2	…	b_s	n_{i+}
a_1	n_{11}	n_{21}	…	n_{1s}	n_{1+}
a_2	n_{21}	n_{22}	…	n_{2s}	n_{2+}
…	…	…	…	…	…
a_r	n_{r1}	n_{r1}	…	n_{rs}	n_{r+}
n_{+j}	n_{+1}	n_{+2}	…	n_{+s}	n

a_i kategorie proměnné A,
b_j kategorie proměnné B,
n_{ij} ($i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, s$) sdružené četnosti absolutní
($p_{ij} = n_{ij}/n$ sdružené četnosti relativní),
n_{i+} marginální četnosti řádkové absolutní ($p_{i+} = n_{i+}/n$ relativní),
n_{+j} marginální četnosti sloupcové absolutní ($p_{+j} = n_{+j}/n$ relativní).

Není-li rozsah výběru n stanoven předem a je dán například dobou realizace šetření, jsou sdružené četnosti n_{ij} nezávislé náhodné veličiny s Poissonovým rozdělením s parametrem m_{ij}, který je jejich střední hodnotou (a tedy očekávanou četností) a také jejich rozptylem. Rovněž rozsah výběru n je v tomto případě náhodná veličina s Poissonovým rozdělením, jeho parametr je $\Sigma\Sigma m_{ij}$.

Je-li rozsah výběru n stanoven, rozdělení každé četnosti n_{ij} v tabulce je binomické s parametry n a π_{ij}, střední hodnotou $m_{ij} = n.\pi_{ij}$ a rozptylem $n.\pi_{ij}(1 - \pi_{ij})$. Četnosti již nejsou vzájemně nezávislé náhodné veličiny (neboť $\Sigma\Sigma n_{ij} = n$) a rozdělení r x s četností v tabulce je multinomické s parametry n, π_{11}, π_{12}, …, π_{rs}, a tedy se středními hodnotami $n\pi_{ij}$, rozptyly $n\pi_{ij}$ (1-π_{ij}) a kovariancemi $-n\pi_{ij}\pi_{i'j'}$. Lze ukázat (viz např. [3]), že uvedené dvě situace se shodují, podmíníme-li sdruženou pravděpodobnostní funkci r x s četností v tabulce s Poissonovým rozdělením součtem $\Sigma\Sigma n_{ij} = n$.

Stanovíme-li předem kromě rozsahu i složení výběru vzhledem k jedné sledované proměnné, tedy fixujme-li například řádkové marginální četnosti n_{i+}, sdružené četnosti v tabulce lze považovat za výsledky třídění r nezávislých náhodných výběrů. Četnosti n_{i1}, n_{i2}, …, n_{is} v jednotlivých řádcích kombinační tabulky pak mají nezávislá multinomická rozdělení (kovariance $C(n_{ij},n_{i'j})$ jsou rovny nule) s parametry n_{i+}, $\pi_{1/i}$, $\pi_{2/i}$, …, $\pi_{s/i}$, ($\pi_{j/i} = \pi_{ij} / \pi_{i+}$). Pokud považujeme za pevně danou kromě marginální řádkové struktury i marginální strukturu

sloupcovou, rozdělení sdružených četností n_{ij} je vícerozměrné hypergeometrické, rozptyly a kovariance se liší o konečnostní násobitel $(n - n_{i+})/(n - 1)$.

Úvahy o rozdělení funkcí výběrových sdružených četností vycházejí nejčastěji z jejich multinomického rozdělení, jeho záměna nezávislými multinomickými či Poissonovým rozdělením totiž nepředstavuje zásadní rozdíl; použití Poissonova rozdělení je v některých případech výhodnější. Předpoklad pevné marginální řádkové i sloupcové struktury pak umožňuje využití exaktních testů, které jsou vhodné právě v řídkých kombinačních tabulkách.

Odchylky od nezávislosti v jednotlivých polích kombinační tabulky shrnuje známá Pearsonova statistika X^2,

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(n_{ij} - m_{ij})^2}{m_{ij}}, \tag{1}$$

vyjadřující stupeň „dobré shody“ mezi sdruženými četnostmi zjištěnými v kontingenční tabulce (n_{ij}) a četnostmi očekávanými v případě nezávislosti veličin (m_{ij}). Rozdělení statistiky X^2 je asymptoticky chí-kvadrát s $(rs - 1)$ stupni volnosti a je často používaným testovým kritériem. Její výpočet ovšem vyžaduje stanovení středních hodnot (očekávaných četností) m_{ij}. Jsou-li proměnné nezávislé, platí $\pi_{ij} = \pi_{i+}\pi_{+j}$ a očekávané četnosti lze tedy odhadnout jako $\hat{m}_{ij} = np_{i+}p_{+j}$, čímž se ovšem počet stupňů volnosti rozdělení statistiky (1) snižuje na $(r - 1)(s - 1)$. Poznamenejme ještě, že v tabulce r x 2 (analogicky 2 x s) lze statistiku (1) zjednodušit na tvar

$$X^2 = \sum_{i=1}^{r}\frac{(n_{i1} - n_{i+}p_{+1})^2}{n_{i+}p_{+1}(1 - p_{+1})} \tag{2}$$

a dále ve čtyřpolní tabulce na tvar

$$X^2 = \frac{n(nn_{11} - n_{1+}n_{+1})^2}{n_{1+}n_{+1}n_{2+}n_{+2}} = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1+}n_{+1}n_{2+}n_{+2}}. \tag{3}$$

Test nezávislosti s užitím výše uvedené statistiky je notoricky znám, stejně tak jako požadavek nutný pro dosažení přijatelné aproximace jejího exaktního rozdělení rozdělením chí-kvadrát, a sice požadavek takového rozsahu výběru, aby očekávané četnosti vesměs dosahovaly hodnoty aspoň 5[1].

V souvislosti s modelováním vztahů mezi veličinami na základě četností v kombinačních tabulkách získalo při testování nezávislosti na významu testové kritérium založené na věrohodnostním poměru. Testové kritérium

$$G^2 = 2\sum_{i}^{r}\sum_{j}^{s} n_{ij}\ln\frac{n_{ij}}{\hat{m}_{ij}} = 2\sum_{i}^{r}\sum_{j}^{s} n_{ij}\ln\frac{n_{ij}.n}{n_{i+}.n_{+j}} \tag{4}$$

(*deviance*) má rovněž asymptoticky chí-kvadrát rozdělení s $(r - 1)(s - 1)$ stupni volnosti. Lze ukázat (viz např. [2]), že pro tabulku o daných rozměrech rozdíl mezi oběma statistikami

[1] Praktické obtíže dodržení této podmínky motivovaly řadu studií, jež vedly ke zjištění, že statistika X^2 je relativně robustní, pokud jde o velikost očekávaných četností, nicméně menších než 5 by mělo být maximálně 20 % z nich (a každá by v takovém případě měla být alespoň jednotková).

konverguje s růstem *n* k nule. Na nedodržení podmínky dostatečně velkých očekávaných četností je ovšem *deviance* citlivější.

Nevyhovující aproximace rozdělení statistik X^2 a G^2 při ověřování nezávislosti v řídkých tabulkách vede k úvahám o jiných použitelných statistikách. V literatuře se lze setkat s různými modifikacemi (Neymanova statistika, Kullbackovou informační statistika, Freemanova-Tukeyova statistika). T. Read a N. Cressie (například [4]) však ukázali, že všechny tyto statistiky jsou jednoho typu (*power divergence statistics*)

$$2nI^{\lambda} = \frac{2}{\lambda(\lambda+1)} \sum_i \sum_j n_{ij} \left[\left(\frac{n_{ij}}{m_{ij}} \right)^{\lambda} - 1 \right], \quad -\infty < \lambda < \infty, \tag{5}$$

a jejich rozdělení je vesměs asymptoticky chí-kvadrát s (*rs* – 1) stupni volnosti, resp. *(r* – 1) (*s* – 1) stupni volnosti, odhadujeme-li m_{ij}. Výraz (5) není definován pro $\lambda = 0$ a $\lambda = -1$, avšak pro $\lambda \to 0$ konverguje ke statistice G^2, pro $\lambda \to -1$ ke Kullbackově informační statistice. Pearsonovu statistiku obdržíme, pokud $\lambda = 1$. Chí-kvadrát aproximace rozdělení různých statistik tohoto typu je na nedostatečné obsazení polí v kontingenční tabulce různě citlivá. Read a Cressie doporučují pro dosažení nejlepší aproximace volit $\lambda = 2/3$.

Důsledkem relativně malého rozsahu výběru při větším počtu kategorií jedné či obou proměnných mohou být málo obsazená či prázdná pole v kontingenční tabulce (tzv. řídké tabulky). Nulové četnosti jsou zde důsledkem toho, že v souboru o malém rozsahu se určitá kombinace kategorií nevyskytla, tedy ne toho, že taková kombinace nemůže existovat. Pro tato prázdná pole platí $m_{ij} > 0$. (Tabulky, které obsahují logicky neobsazená pole, tzv. nekompletní kontingenční tabulky, kdy $m_{ij} = 0$ a nutně také $\hat{m}_{ij} = 0$, vyžadují zvláštní režim zacházení a v tomto textu se jimi nezabývám.)

V případě malého rozsahu výběru, kdy jsou některá pole v tabulce nedostatečně obsazena či nejsou obsazena vůbec (tzv. řídké kontingenční tabulky), nelze považovat aproximaci rozdělení výše uvedených statistik za vyhovující. V literatuře často doporučovaným řešením je zmenšení počtu kategorií proměnných jejich slučováním. Kromě věcných námitek je však třeba rovněž uvážit, že takový postup může vést k jinému zjištění, než původní tabulka.

Vytvoříme-li totiž v kontingenční tabulce *r* x *s* celkem *k* takových tabulek o menších rozměrech, které mají uvedené vlastnosti, pak statistiky G_1^2, G_2^2, …, G_k^2 z nich spočtené jsou nezávislé náhodné veličiny s asymptoticky chí-kvadrát rozdělením, pro něž platí $G^2 = G_1^2 + G_2^2 + ... + G_k^2$. Pro parametry jejich rozdělení pak platí $df = df_1 + df_2 + \ldots + df_k$. Maximální počet dílčích tabulek je v takovém případě $k = (r-1)(s-1)$, kdy všechny dílčí tabulky jsou čtyřpolní a lze je vytvořit například takto:

$\sum_{a<i} \sum_{b<j} n_{ab}$	$\sum_{a<i} n_{aj}$
$\sum_{b<j} n_{ib}$	n_{ij}

Všechna chí-kvadrát rozdělení pak mají jeden stupeň volnosti. Pro statistiky X^2 platí obdobný vztah jako pro G^2 asymptoticky.

Tabulka získaná slučováním řádků, příp. sloupců původní tabulky je vůči ní tabulkou dílčí a testová kritéria tedy vždy budou nižší než v původní tabulce. Nižší je ovšem také počet stupňů volnosti jejich rozdělení. Tím může dojít k rozdílným výsledkům testů.

Zajímavější možností při testování nezávislosti v řídkých kombinačních tabulkách je použití exaktních testů. Relativně známý je v této souvislosti Fisherův (faktoriálový) test pro čtyřpolní tabulku ($r = s = 2$). Považujeme-li marginální četnosti v takové tabulce za pevně dané, jsou všechny sdružené četnosti určeny jedinou z nich, např. n_{11}. Její rozdělení je hypergeometrické se střední hodnotou $n_{1+}n_{+1}/n$ a s rozptylem

$$\frac{n_{1+}n_{2+}n_{+1}n_{+2}}{n^2(n-1)}.$$

V soustavě všech možných tabulek s pevnými marginálními četnostmi a s různým uspořádáním sdružených četností (neboli s určitou hodnotou četnosti n_{11}) lze identifikovat takové, které by hypotéze o nezávislosti byly ještě méně příznivé, než rozmístění četností v tabulce s výběrovými daty. Označíme-li menší z dvojice n_{1+} a n_{+1} jako *min*, pak máme na mysli tabulky, v nichž n_{11} menší než *min*/2 klesá postupně k nule, nebo n_{11} větší než *min*/2 roste postupně k *min*. Součet pravděpodobností pořízení všech takových tabulek (které také klesají k nule) pak rozhoduje o výsledku testu nezávislosti, neboť dává minimální hladinu významnosti, na níž by bylo možno hypotézu o nezávislosti zamítnout.

Použití exaktního testu obvykle představuje značný objem výpočtů. Ve čtyřpolní tabulce bývalo proto často v literatuře doporučováno použití korigované Pearsonovy statistiky (tzv. Yatesova korekce)

$$X_C^2 = \frac{n(\left|n_{11}n_{22} - n_{12}n_{21}\right| - \frac{n}{2})^2}{n_{1+}n_{+1}n_{2+}n_{+2}} \qquad (6)$$

jako adekvátní aproximace Fisherova exaktního testu. Nejde tedy o lepší aproximaci rozdělení statistiky rozdělením chí-kvadrát, jak bývá často mylně tato korekce chápána. Statistické výpočetní systémy však dnes umožňují použití Fisherova testu i v rozsáhlejších souborech a v minulosti hojně diskutovaná korekce tak poněkud pozbývá na významu.

Jsou-li fixovány marginální četnosti v kontingenční tabulce r x s, $(s - 1)$ sdružených četností v každém řádku a $(r - 1)$ sdružených četností v každém sloupci určuje zbývající. Exaktní testy nezávislosti lze založit podobně jako Fisherův test na výpočtu pravděpodobností vzniku různě uspořádaných tabulek a zjištění, s jakou pravděpodobností vznikne tabulka ještě méně příznivá hypotéze nezávislosti, než je tabulka s výběrovými údaji (Freemanův-Haltonův test). Jinou možností je stanovení hodnot jakékoliv statistiky (například Pearsonovy) a zjištění všech tabulek s hodnotami zvolené statistiky méně příznivými hypotéze o nezávislosti než v tabulce s výběrovými daty spolu s pravděpodobnostmi pořízení takových tabulek. P-hodnota testu pak je dána součtem těchto pravděpodobností.

Se zvětšováním rozsahu datového souboru a rozměrů tabulky však počet jejích možných uspořádání velmi rychle narůstá a zvětšuje se značně objem potřebných výpočtů.. Řešením může být použití metody Monte Carlo, kdy je náhodně generován zvolený (co nejvyšší) počet odpovídajících tabulek a *p*-hodnota testu je odhadnuta na základě podílu tabulek s méně příznivými hodnotami používané statistiky mezi generovanými tabulkami.

Vyvrácení hypotézy o nezávislosti vede k podrobnějšímu zkoumání příčin, tedy především k hledání nejvýznamnějších rozdílů mezi zjištěnými a odhadnutými očekávanými četnostmi v jednotlivých polích kontingenční tabulky – reziduí $(n_{ij} - \hat{m}_{ij})$. Klasická rezidua však nejsou dobrým diagnostickým nástrojem nedostatku dobré shody, neboť samotné odchylky mají bez konfrontace s nějakým kritériem jen malou vypovídací schopnost. Normováním reziduí jejich dělením směrodatnými odchylkami,

$$e_{ij} = \frac{n_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}}}, \tag{7}$$

lze dosáhnout toho, že mají asymptoticky normální rozdělení s nulovou střední hodnotou, jejich variabilita však je většinou menší, než jednotková. Vhodnějším diagnostickým nástrojem nedostatku dobré shody jsou adjustovaná rezidua (Habermanova)

$$e_{ij} = \frac{n_{ij} - \hat{m}_{ij}}{\sqrt{\hat{m}_{ij}(1 - p_{i+})(1 - p_{+j})}}. \tag{8}$$

Jejich rozdělení je asymptoticky normované normální a k vyhodnocení významnosti každé odchylky tedy stačí konfrontace s běžně známými kvantily tohoto rozdělení. Výpočetní systémy (například SPSS) obvykle rezidua tohoto typu počítají a na jejich základě lze tak snadno sestavit oblíbená znaménková schémata znázorňující jedním, dvěma či třemi znaménky (plus či minus) v jednotlivých polích tabulky odchylky od nezávislosti významné na hladině významnosti 5 %, 1 %, resp. 0,1 %.

Literatura:

[1] Agresti, A.: Categorical Data Analysis, John Wiley α Sons, 1990
[2] Kendall, M.G. – Stuart, A.: The Advanced Theory of Statistics, Vol.2: Inference and Relationship, Ch. Griffin & Comp., London 1967
[3] Pecáková, I.: Analýza a modelování souvislostí kategoriálních proměnných, habilitační práce, Praha 2004
[4] Read, T. R. C., Cressie, N. A. C.: Goodness-of-fit statistics for discrete multivariate data, Springer 1988

Doc. Ing. Iva Pecáková, CSc.
The University of Economics
Faculty of Informatics and Probability
Department of Statistics and Probability
Prague, Czech Republic
e-mail: pecakova@vse.cz

Testy pro alternativní proměnné ve statistických programových systémech*

Hana Řezanková

Abstract. The aim of the paper is to compare results obtained by binomial and McNemar's tests in statistical packages SAS, SPSS and STATGRAPHICS Plus. The differences consist in using one-sided or two-sided alternative hypotheses, exact or asymptotic tests, and distinct formulas for asymptotic statistics. Binomial exact test and approximations by normal and chi-square distributions are used.

1. Úvod

Statistické programové systémy obvykle zahrnují stejné základní statistické nástroje pro testování hypotéz. Samozřejmou součástí jsou například t testy o středních hodnotách normálního rozdělení v členění na jednovýběrové a dvouvýběrové (pro 2 nezávislé a 2 závislé výběry, tj. párový t test). Též výsledky můžeme očekávat stejné.

Některé systémy již však nezahrnují testy o parametru π binomického rozdělení. Pokud máme programové produkty, v nichž jsou tyto testy k dispozici, nemusíme dostat stejné výsledky. Navíc se systémy mohou lišit z hlediska uvažovaných alternativních hypotéz.

Tento příspěvek porovnává postupy používané při binomickém testu v programových systémech SAS, SPSS a STATGRAPHICS a při McNemarově testu v systémech SAS a SPSS. Jde tedy jednak o jednovýběrový test o podílu (relativní četnosti), jednak o párový test o shodě podílů.

2. Binomický a McNemarův test

Při **binomickém testu** vycházíme ze zjištěných četností dvou kategorií dichotomické (alternativní) proměnné. Označme si rozsah výběrového souboru (počet sledovaných statistických jednotek) jako n, zjištěné absolutní četnosti jako n_1 a n_2, relativní jako p_1 a p_2. Testujme hypotézu H_0: $\pi_1 = \pi_{1,0}$, kde $\pi_{1,0}$ je očekávaná relativní četnost v základním souboru.

Můžeme použít buď exaktní text s využitím binomického rozdělení, nebo aproximaci normovaným normálním rozdělením. Jako podmínka pro využití této aproximace se v literatuře uvádí, aby součin $n\pi_{1,0}(1 - \pi_{1,0})$ byl větší než hodnota 9.

V případě, že $\pi_{1,0} \geq p_1$, je jednostranná alternativní hypotéza stanovena jako levostranná, tj. H_1: $\pi_1 < \pi_{1,0}$. Minimální hladina významnosti, od které zamítáme nulovou hypotézu vůči této alternativní hypotéze, se pro $n_1 < n_2$ exaktně počítá podle vztahu

$$\alpha' = \sum_{i=0}^{n_1} \binom{n}{i} (\pi_{1,0})^i (1 - \pi_{1,0})^{n-i}.$$

Jde tedy o výpočet pravděpodobnosti, že náhodná veličina s binomickým rozdělením nabude hodnoty n_1 nebo menší, tzn. o výpočet distribuční funkce v bodě n_1, což budeme značit jako $F(n_1)$. Přitom pravděpodobnost, že nastane sledovaný náhodný jev, je $\pi_{1,0}$ a počet náhodných pokusů je n.

Pro $\pi_{1,0} < p_1$ je jednostranná alternativní hypotéza stanovena jako pravostranná, tj. H_1: $\pi_1 > \pi_{1,0}$. Minimální hladina významnosti, od níž zamítáme nulovou hypotézu vůči této alternativní hypotéze, se pro $n_1 < n_2$ exaktně počítá podle vztahu $\alpha' = 1 - F(n_1) + P(n_1)$, kde $P(n_1)$ je pravděpodobnost, že náhodná veličina nabude hodnoty n_1. Jde tedy o výpočet pravděpodobnosti, že náhodná veličina s binomickým rozdělením nabude hodnoty n_1 nebo větší, což lze též zapsat jako $\alpha' = 1 - F(n_1 - 1)$.

V případě, kdy $\pi_{1,0} = 0{,}5$, se testuje nulová hypotéza o shodě podílů. Můžeme ji zapsat jako H_0: $\pi_1 = \pi_2$, neboť platí, že $\pi_1 + \pi_2 = 1$. Obvykle se testuje vůči oboustranné alternativní hypotéze H_1: $\pi_1 \neq \pi_2$. Za předpokladu, že $n_1 < n_2$, je minimální hladina významnosti, od které zamítáme nulovou hypotézu, dána vztahem $\alpha' = 2 \cdot F(n_1)$.

* Tento příspěvek byl vytvořen za podpory grantu GAČR 201/05/0079.

Při aproximaci normovaným normálním rozdělením se zjišťuje hodnota normované veličiny. Výpočet se provede tak, že se od zjištěné četnosti sledované kategorie odečte očekávaná četnost a výsledný rozdíl se dělí směrodatnou odchylkou. Některé programové systémy používají korekce, a to tak, že přičítají (resp. odečítají) hodnotu 0,5, viz [7].

McNemarův test je párový test o shodě podílů, který může být využit například při zjišťování, zda se shodují nebo liší názory respondentů ve dvou různých obdobích. V programových systémech bývá zařazován buď k neparametrickým testům, nebo k testům na základě dat z kontingenční tabulky.

Uvažujme dvě alternativní proměnné a označme sdružené četnosti ve čtyřpolní tabulce jako n_{ij}, kde i a j nabývají hodnot 1 a 2 (pořadí kategorie příslušné alternativní proměnné). Při McNemarově testu se testuje nulová hypotéza o shodě četností v políčkách na vedlejší diagonále. Hypotézy můžeme zapsat ve tvaru H_0: $\pi_{12} = \pi_{21}$, H_1: $\pi_{12} \neq \pi_{21}$.

V případě exaktního testu je minimální hladina významnosti pro zamítnutí nulové hypotézy daná vztahem

$$\alpha' = 2 \sum_{i=0}^{\min\{n_{12}, n_{21}\}} \binom{n_{12} + n_{21}}{i} (0{,}5)^{n_{12}+n_{21}} . \quad (1)$$

Polovina této hodnoty vyjadřuje pravděpodobnost, že náhodná veličina s binomickým rozdělením (pravděpodobnost, že nastane sledovaný náhodný jev, je 0,5 a počet náhodných pokusů je $n_{12} + n_{21}$) nabude hodnoty $\min\{n_{12}, n_{21}\}$ nebo menší.

Při větším počtu náhodných pokusů lze využít aproximaci binomického rozdělení rozdělením chí-kvadrát. Součet četností ($n_{12} + n_{21}$) by měl být aspoň 10, aby byla zajištěna podmínka pro aplikovaný chí-kvadrát test dobré shody, kdy jsou jednotlivé zjištěné četnosti porovnávány s očekávanými počítanými podle vztahu $(n_{12} + n_{21})/2$ (v [2] se uvádí více než 10, v [4] aspoň 30). Úpravou vzorce pro statistiku chí-kvadrát pro dvě kategorie získáváme vztah

$$Q_M = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} . \quad (2)$$

Za předpokladu platnosti nulové hypotézy má tato náhodná veličina chí-kvadrát rozdělení s jedním stupněm volnosti. Některé programové systémy používají korekci, a to tak, že od absolutní hodnoty rozdílu hodnot v čitateli odečítají hodnotu 1, viz vzorec (6).

3. Postupy použité v programových systémech

Výsledky dále uvedené byly získány pomocí systémů *SAS Learning Edition 2.0*, který využívá uživatelské rozhraní *Enterprise Guide*, *SPSS 11. 5 for Windows* a *STATGRAPHICS Plus for Windows*, verze 3.1.

Pokud jde o **binomický test**, pak systém SAS uvádí meze oboustranného intervalu spolehlivosti a výsledky (minimální hladinu významnosti pro zamítnutí nulové hypotézy) jak pro jednostrannou, tak pro oboustrannou alternativní hypotézu. Lze zobrazit výsledky jak pro exaktní test, tak pro aproximaci normálním rozdělením, a to bez ohledu na rozsah výběrového souboru. Při aproximaci normálním rozdělením se počítá hodnota Z podle vztahu

$$Z = \frac{\min\{n_1, n_2\} - n\pi}{\sqrt{n\pi(1-\pi)}} , \quad (3)$$

kde hodnota π je stanovena následujícím způsobem. Pro $\min\{n_1, n_2\} = n_1$ je $\pi = \pi_{1,0}$, v opačném případě $\pi = (1 - \pi_{1,0})$. Při testování vůči levostranné alternativní hypotéze jde o kvantil $u_{\alpha'}$, kde α' je minimální hladina významnosti, od které zamítáme nulovou hypotézu. Platí tedy, že $\alpha' = \Phi(Z)$, kde $\Phi(Z)$ je hodnota distribuční funkce normovaného normálního rozdělení v bodě Z. Pro $\pi_{1,0} = 0{,}5$ je navíc možné jako alternativu použít chí-kvadrát test dobré shody, kdy jsou použity vzorce (1) a (2).

Systém SPSS meze intervalu spolehlivosti neuvádí. Pokud $\pi_{1,0} \neq 0{,}5$, zobrazují se výsledky pro jednostrannou alternativní hypotézu, je-li $\pi_{1,0} = 0{,}5$, výsledky pro oboustrannou alternativní hypotézu. Pro $n_1 + n_2 \leq 25$ systém uvádí, že se používá binomické rozdělení, pro $n_1 + n_2 > 25$ se uvádí, že byla použita aproximace normálním rozdělením. Získané výsledky však odpovídají hodnotám vypočítaným

na základě distribuční funkce binomického rozdělení. Přibližně stejné výsledky bychom získali pomocí korigované hodnoty (při testování vůči levostranné alternativní hypotéze jde o kvantil $u_{\alpha'}$)

$$Z = \frac{\min\{n_1, n_2\} + 0{,}5 - n\pi}{\sqrt{n\pi(1-\pi)}}. \quad (4)$$

Při testování vůči oboustranné alternativní hypotéze jde o kvantil $u_{\alpha'/2}$, kde α' je minimální hladina významnosti, od které zamítáme nulovou hypotézu o shodě podílů u daných dvou kategorií. Tuto hladinu významnosti tedy spočteme jako $\alpha' = 2 \cdot \Phi(Z)$.

Při testování vůči pravostranné alternativní hypotéze bychom mohli minimální hladinu významnosti vypočítat na základě vztahu $\alpha' = 1 - \Phi(Z)$, přičemž se uvažuje četnost $\min\{n_1, n_2\} - 1$, tj.

$$Z = \frac{\min\{n_1, n_2\} - 0{,}5 - n\pi}{\sqrt{n\pi(1-\pi)}}. \quad (5)$$

STATGRAPHICS uvádí meze oboustranného intervalu spolehlivosti a výsledky pouze pro oboustrannou alternativní hypotézu. Tyto výsledky jsou získány s použitím binomického rozdělení.

V případě **McNemarova testu** je v systému SAS použito testové kritérium (2). Test je nabízen jako míra souhlasu v rámci analýzy kontingenčních tabulek.

V SPSS lze tento test provést jednak v rámci analýzy kontingenčních tabulek, jednak v rámci neparametrických párových testů. V prvním případě se používá exaktní test, kdy se minimální hladina významnosti zjišťuje podle vztahu (1), a to bez ohledu na rozsah četností. Ve druhém případě se postup liší podle četností. Pro $n_{12} + n_{21} \leq 25$ se používá exaktní test, pro $n_{12} + n_{21} > 25$ pak aproximace rozdělením chí-kvadrát. Testové kritérium s opravou na spojitost, jehož hodnotu porovnáváme s kvantilem z chí-kvadrát rozdělení s jedním stupněm volnosti, je dáno vztahem

$$Q_M = \frac{\left(|n_{12} - n_{21}| - 1\right)^2}{n_{12} + n_{21}}. \quad (6)$$

V systému STATGRAPHICS není McNemarův test k dispozici.

4. Porovnání výsledků

V této části budou uvedeny výsledky čtyř příkladů získané pomocí jednotlivých programových systémů. Budeme testovat podíl pro malý a velký výběrový soubor a shodu podílů v případě závislých výběrů, v členění na malý a velký rozsah četností.

Příklad 1 – binomický test pro malý výběr

Porovnejme výsledky získané na základě výběrového souboru o rozsahu $n = 5$, přičemž byly zjištěny četnosti $n_1 = 1$ a $n_2 = 4$ (z pěti respondentů pouze jeden uvedl zájem o určitou problematiku). Budeme uvažovat tři různé nulové hypotézy, a to $\pi_1 = 0{,}5$, $\pi_1 = 0{,}4$, $\pi_1 = 0{,}1$. Získané minimální hladiny významnosti pro zamítnutí nulové hypotézy jsou uvedeny v tabulce 1.

Pro srovnání jsou též uvedeny výsledky, kdy byla hodnota testového kritéria spočtena „ručně“ podle vzorce (3) a hladiny významnosti získány jako $\alpha' = \Phi(Z)$ u levostranné alternativní hypotézy a $\alpha' = 1 - \Phi(Z)$ u alternativní hypotézy pravostranné.

Příklad 2 – binomický test pro velký výběr

Porovnejme výsledky získané na základě výběrového souboru o rozsahu $n = 50$, přičemž byly zjištěny četnosti $n_1 = 17$ a $n_2 = 33$ (z 50 respondentů jich 17 uvedlo zájem o určitou problematiku). Budeme uvažovat tři různé nulové hypotézy, a to $\pi_1 = 0{,}5$, $\pi_1 = 0{,}4$, $\pi_1 = 0{,}3$. Získané minimální hladiny významnosti pro zamítnutí nulové hypotézy jsou uvedeny v tabulce 2. Zařazen je též výsledek získaný pomocí asymptotického chí-kvadrát testu dobré shody v systému SAS.

Pro srovnání jsou dále uvedeny výsledky, kdy byla hodnota testového kritéria spočtena „ručně“ podle vzorců (3), (4), resp. (5), a hladiny významnosti získány výše popsanými postupy. Ruční výpočet byl rovněž proveden pro chí-kvadrát test dobré shody pomocí vzorce (6) s četnostmi n_1 a n_2.

Tabulka 1. Minimální hladiny významnosti pro zamítnutí nulové hypotézy pro příklad 1

Systém	Metoda	Alt. hypotéza	Nulová hypotéza		
			$\pi_1 = 0,5$	$\pi_1 = 0,4$	$\pi_1 = 0,1$
SAS	binom. rozd.	jednostranná	0,1875	0,3370	0,4095
	Norm. rozd.	jednostranná	0,0899	0,1807	0,2280
SPSS			0,375*)	0,337**)	0,410**)
STATGRAPHICS			0,375*)	0,6739*)	0,819*)
ruční výpočty	vzorec (3)	jednostranná	0,0898	0,1806	0,2281

*) oboustranná alternativní hypotéza
**) jednostranná alternativní hypotéza

Tabulka 2. Minimální hladiny významnosti pro zamítnutí nulové hypotézy pro příklad 2

Systém	Metoda	Alt. hypotéza	Nulová hypotéza		
			$\pi_1 = 0,5$	$\pi_1 = 0,4$	$\pi_1 = 0,3$
SAS	binom. rozd.	jednostranná	0,0164	0,2369	0,3161
	Norm. rozd.	jednostranná	0,0118	0,1932	0,2685
	chí-kvadrát	oboustranná	0,0237	x	x
SPSS			0,033*)	0,237**)	0,316**)
STATGRAPHICS			0,033*)	0,4738*)	0,6322*)
ruční výpočty	vzorec (3)	jednostranná	0,0118	0,1932	0,2686
	vzorce (4), (5)	jednostranná	0,01696	0,2351	0,3217
	vzorec (6)	oboustranná	0,0339	x	x

*) oboustranná alternativní hypotéza
**) jednostranná alternativní hypotéza

Příklad 3 – McNemarův test pro malý rozsah četností

Proveďme test pro případ, kdy na základě dvou alternativních proměnných byly zjištěny sdružené četnosti $n_{12} = 1$ a $n_{21} = 4$. Získané minimální hladiny významnosti pro zamítnutí nulové hypotézy jsou uvedeny v tabulce 3.

Tabulka 3. Minimální hladiny významnosti pro zamítnutí nulové hypotézy pro příklad 3

Systém	Metoda	α'
SAS	chí-kvadrát	0,1797
SPSS	binom. rozd.	0,375

Příklad 4 – McNemarův test pro velký rozsah četností

Proveďme test pro případ, kdy na základě dvou alternativních proměnných byly zjištěny sdružené četnosti $n_{12} = 12$ a $n_{21} = 16$. Získané minimální hladiny významnosti pro zamítnutí nulové hypotézy jsou uvedeny v tabulce 4.

Tabulka 4. Minimální hladiny významnosti pro zamítnutí nulové hypotézy pro příklad 4

Systém	Metoda	α'
SAS	chí-kvadrát	0,4497
SPSS	binom. rozd.	0,572
	chí-kvadrát	0,571

5. Závěr

Námětů pro porovnání statistických postupů v různých programových systémech bychom našli velké množství. Ani výše uvedené srovnání jedné dílčí problematiky není úplné. Rozdíly existují také ve způsobu zadání vstupu pro testování hypotéz. Systémy SAS a SPSS vycházejí při binomickém testu ze zjištěných hodnot proměnných, které mohou být zadány jednotlivě, nebo pomocí četností jednotlivých variant hodnot (tyto četnosti jsou specifikovány při zadání testu, nebo před zadáním jako váhy). STATGRAPHICS provádí test na základě zadaných parametrů, kterými jsou $\pi_{1,0}$, p_1 a n.

Z porovnání uvedeného v tomto příspěvku vyplývá, že v případě binomického testu poskytují všechny tři programové systémy výsledky získané pomocí exaktního testu. Protože minimální hladina významnosti pro zamítnutí nulové hypotézy je pro oboustrannou alternativní hypotézu dvojnásobkem minimální hladiny významnosti pro uvažovanou alternativní hypotézu jednostrannou, lze při absenci jedné z možností druhou snadno dopočítat. Asymptotický test nabízejí systémy SAS a SPSS, přičemž SAS ponechává výběr na uživateli, zatímco SPSS zohledňuje velikost výběru. Pro oboustrannou alternativní hypotézu navíc systém SAS nabízí chí-kvadrát test dobré shody doplněný o exaktní výsledky získané pomocí binomického rozdělení. Jde však o duplicitní výpočty, neboť exaktní test je shodný s binomickým testem a aproximace chí-kvadrát testem dává shodné výsledky jako aproximace normovaným normálním rozdělením (viz též [6]).

Pokud jde o McNemarův test, systém SAS vychází ze statistiky používané při chí-kvadrát testu dobré shody. Není zahrnuta exaktní varianta pro malý rozsah četností. Navíc se ani nezobrazuje upozornění na nízké obsazení četností v políčkách kontingenční tabulky, i když při testování podílu pro jeden výběr se takové upozornění vypisuje. V systému SPSS se výsledky mohou lišit podle toho, zda je test proveden v rámci analýzy kontingenčních tabulek, nebo v rámci neparametrických párových testů. Při aproximaci chí-kvadrát rozdělením se používá statistika korigovaná na spojitost.

Literatura

[1] Arltová, M., Bílková, D., Jarošová, E., Pourová, Z.: *Příklady k předmětu Statistika A, 2. vydání*. Oeconomica, Praha 2004.

[2] Blatná, D.: *Statistické aspekty terénních průzkumů II*. VŠE, Praha 1994.

[3] Hindls, R., Hronová, S., Seger, J.: *Statistika pro ekonomy*. Professional Publishing, Praha 2003.

[4] Marek, L. a kol.: *Statistika pro ekonomy – aplikace*. Professional Publishing, Praha 2005.

[5] Pecáková, I.: *Statistické aspekty terénních průzkumů I*. VŠE, Praha 1995.

[6] Pecáková, I., Novák, I., Herzmann, J.: *Pořizování a vyhodnocování dat ve výzkumech veřejného mínění*. Oeconomica, Praha 2004.

[7] Řezanková, H.: *Analýza kategoriálních dat*. Oeconomica, Praha 2005.

Adresa autorky: doc. Ing. Hana Řezanková, CSc., katedra statistiky a pravděpodobnosti, Vysoká škola ekonomická v Praze, nám. W. Churchilla 4, 130 67 Praha 3, Česká republika, e-mail: rezanka@vse.cz

Analýza časových řad míry nezaměstnanosti dvou regionů Ústeckého kraje

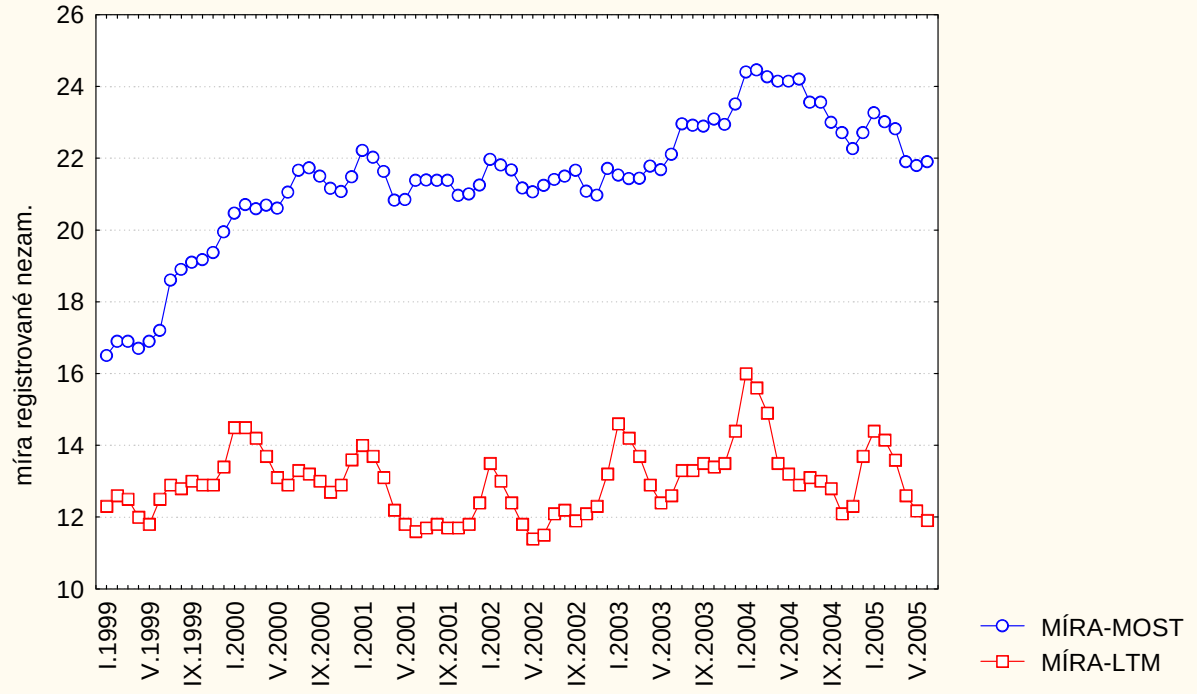
Jana Šimsová

Úvod

Ústecký kraj se potýká s vysokou nezaměstnaností a je z hlediska sledování míry nezaměstnanosti rozdělen do 7 regionů: Děčínsko, Teplicko, Chomutovsko, Lounsko, Litoměřicko, Ústecko, Mostecko. (Tyto regiony odpovídají dřívějším okresům). Nejvyšší nezaměstnanost Ústeckého kraje dlouhodobě vykazuje mostecký region. Tento region se nachází v pánevní oblasti a problémy s vysokou nezaměstnaností jsou zde spojeny s dřívější úzkou specializací tohoto regionu na těžební průmysl, který v posledních letech zaznamenává útlum, i s nepříznivým stavem kvalifikační struktury. Litoměřicko je region ležící převážně v Polabské nížině a je tedy regionem, ve kterém zemědělství hraje značnou roli.

Data byla zpracovávána pomocí SW STATISTICA v rámci grantu GAČR 402/04/0263 „Statistické metody a nezaměstnanost“. V příspěvku jsou analyzovány časové řady míry registrované nezaměstnanosti. Tato míra registrované nezaměstnanosti (dále jen míra nezaměstnanosti) je vyjadřována v % a počítána jako podíl uchazečů o zaměstnání registrovaných na úřadech práce ku počtu ekonomicky aktivních obyvatel. Tento počet ekonomicky aktivních obyvatel je součet počtu zaměstnaných (získaný z výběrových šetření pracovních sil prováděných statistickými úřady) a počtu uchazečů o zaměstnání evidovaných na úřadech práce. Od třetího čtvrtletí 2004 došlo ke změně metodiky výpočtu míry nezaměstnanosti (mezi ekonomicky aktivní obyvatelstvo se připočte ještě počet pracujících cizinců podle evidence ministerstva práce a sociálních věcí), aby se vyhovělo požadavkům Eurostatu a hodnoty míry nezaměstnanosti mohly být porovnávány s hodnotami míry nezaměstnanosti v ostatních zemích Evropské unie. V příslušných časových řadách jsou však uvedeny hodnoty počítané podle staré metodiky, aby nedošlo ke zkreslení. Obě časové řady jsou měsíční, začínají údajem z ledna 1999 a končí údajem z června 2005. Zkoumané řady obsahují tedy 78 hodnot, které představují míru nezaměstnanosti v % na Mostecku a Litoměřicku. Vývoj míry nezaměstnanosti v obou sledovaných regionech v uvažovaném období ukazuje graf č.1. Z tohoto grafu je zřetelné, že s větší mírou nezaměstnanosti se potýká Mostecko. Obě časové řady vykazují sezónní chování.

Graf 1

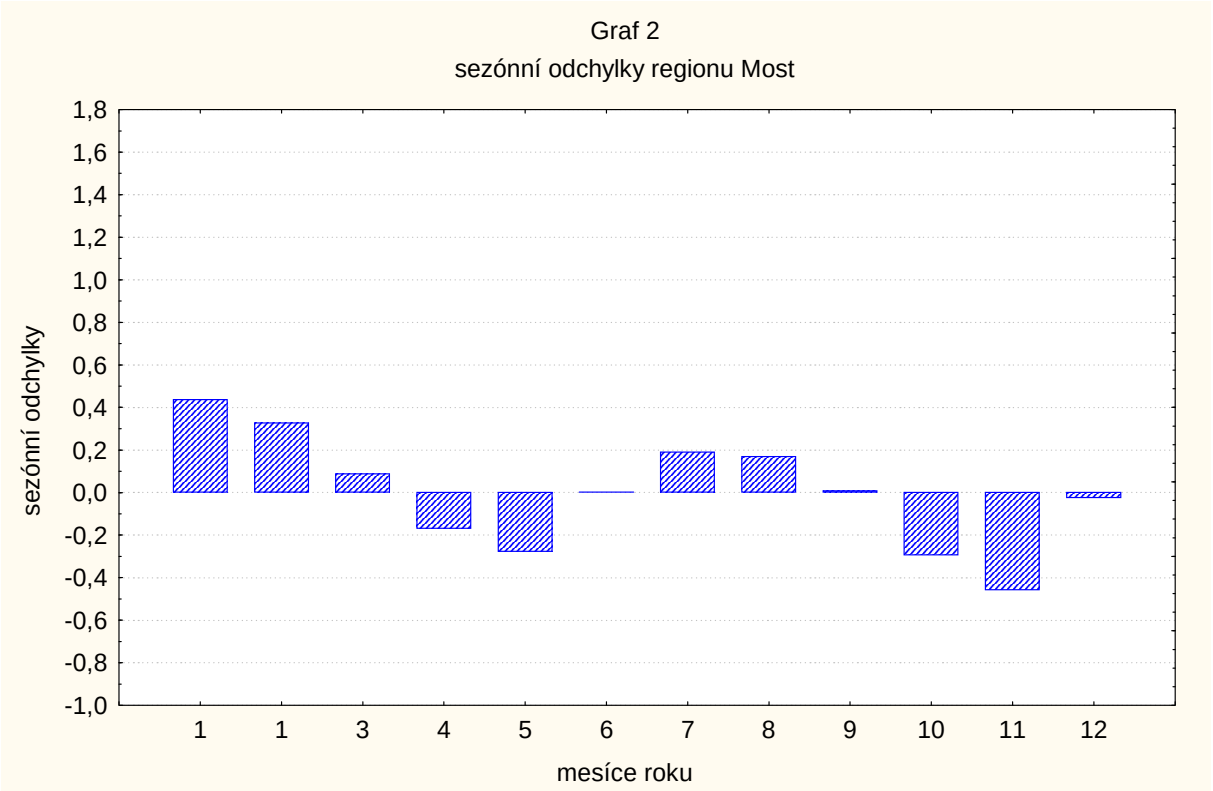


Zdroj:vlastní zpracování

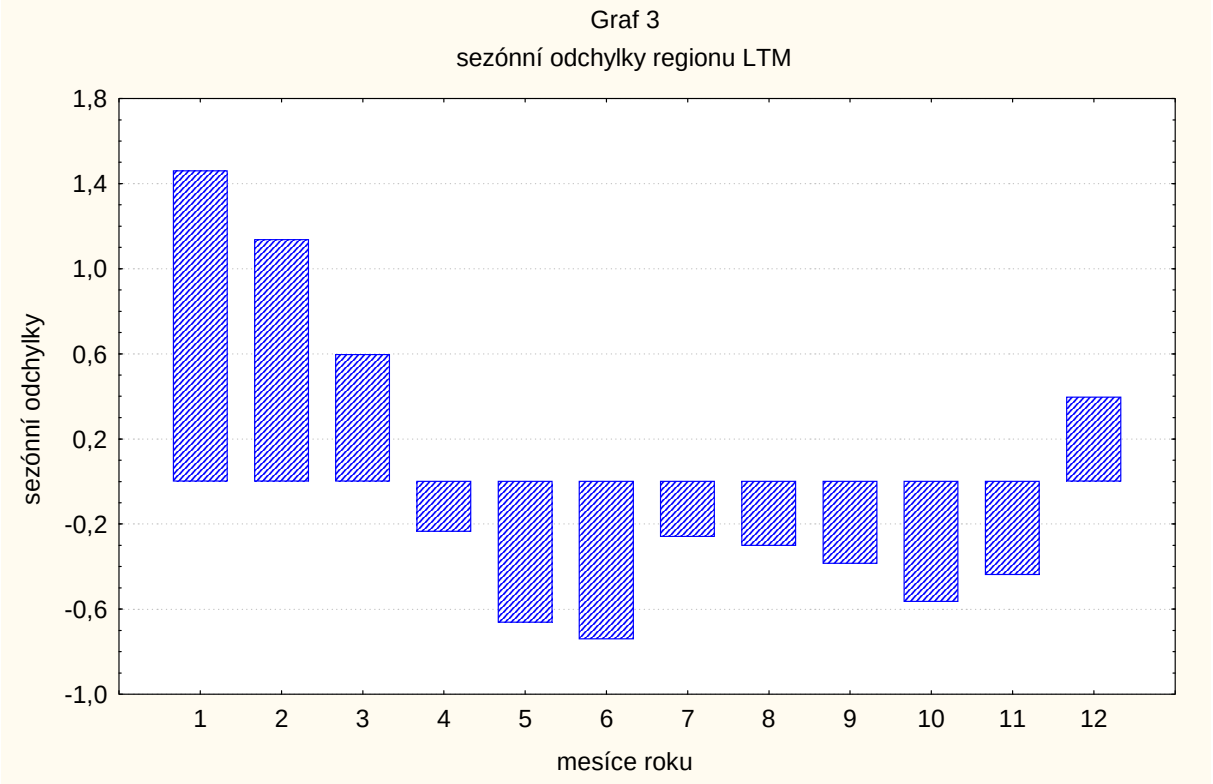
Sezónnost

K popisu obou časových řad byl zvolen aditivní model a k posouzení sezónní složky byly spočteny normované sezónní odchylky a to tak, že původní řada byla vyhlazena centrovanými klouzavými průměry délky 12 a rozdíly původní časové řady a vyhlazené řady byly pro každý měsíc zprůměrovány. Po vynormování jsme dostali sezónní odchylky. Tyto sezónní odchylky pro oba regiony jsou znázorněny na grafech č.2 a č.3 .

Graf 2



Graf 3



Zdroj: výpočty autora

Z těchto grafů vyplývá, že výrazněji se sezónní chování projevuje na Litoměřicku. Toto není nijak překvapující, uvědomíme-li si, že na tuto složku časové řady míry nezaměstnanosti mají vliv sezónní práce nejen ve stavebnictví, ale i v zemědělství a Litoměřicko je převážně zemědělský region. Z grafů je zřejmé, že v jarních měsících (duben, květen, červen) dochází pravidelně k poklesu míry nezaměstnanosti v důsledku sezónních prací. Na Mostecku dochází opět ke zvýšení míry nezaměstnanosti v měsících červnu, červenci, srpnu a září. Tento sezónní výkyv nastává v důsledku konce školního roku a ukončení odvolacích řízení na vysokých školách, kdy jsou do evidence úřadů práce zařazováni absolventi, kteří nenaleznou práci nebo byli neúspěšní v přijímacím řízení na vysokých školách. Na Litoměřicku se tento sezónní výkyv projevuje pouze nižším poklesem míry nezaměstnanosti. Oba regiony zaznamenávají nejvyšší sezónní nárůst míry nezaměstnanosti v zimních měsících, tedy na přelomu roku, kdy dochází k ukončení sezónních prací, ale taky veřejně prospěšných prací, tedy míst zřízených úřady práce v rámci aktivní politiky nezaměstnanosti.

Dekompoziční extrapolační analýza

Dlouhodobá tendence vývoje u obou časových řad byla modelována různými funkcemi a u každé modelové funkce byla spočtena střední kvadratická chyba a index determinace v %. Výsledky jsou uvedeny v tabulce č.1.

Tabulka č.1.

Region	**Trendová funkce**	**Funkční předpis**	**Střední kvadratická chyba**	**Index determinace**
Most	Přímka	y=18,8+0,065x	1,08	66,82%
	Parabola	$y=17,46697+0,16618x-0,001274x^2$	0,75	77,07%
	Logaritmická funkce	y=15,3+4,128log(x)	0,59	81,77%
Litoměřice	Přímka	y=12,67+0,008x	0,87	3,28%
	Polynom čtvrtého stupně	$y=11,36+0,38x0,023x^2+0,00046x^3-0,000003x^4$	0,58	35,71%
	Logaritmická funkce	y=12,39+0,39log(x)	0,88	2,66%

Zdroj: výpočty autora

Z této tabulky vyplývá, že pro region Most nejlépe popisuje trendovou složku této časové řady logaritmická funkce, její střední kvadratická chyba je nejnižší a index determinace je nejvyšší. U časové řady litoměřického regionu je situace daleko složitější. Z nabídnutých funkcí nejlépe popisuje časovou řadu polynom čtvrtého stupně. Její střední kvadratická chyba je nejnižší a index determinace nejvyšší. Ale tato trendová funkce je velice nevhodná pro extrapolaci, pro následující období totiž prudce klesá. Proto pro popis dlouhodobé tendence trendové složky této časové řady byla zvolena přímka. Prognózu na období nejbližších dvanácti měsíců (tedy od července 2005 do června 2006) provedeme tak, že pro toto období spočteme hodnoty zvolené trendové funkce a přičteme k ní sezónní odchylky. Hodnoty předpovědí pro obě časové řady touto metodou jsou uvedeny ve třetím sloupci tabulky č.2.

Spektrální analýza

Spektrální analýza [2] je metoda, která předpokládá, že časovou řadu s konstantním trendem můžeme rozložit do funkcí sinus a cosinus různých frekvencí. Tedy můžeme psát

$$y_t = \frac{a_0}{2} + \sum_{j=1}^{H} a_j \cos \omega_j t + \sum_{j=1}^{H} b_j \sin \omega_j t + \varepsilon_t, \qquad t = 1,2,.....,n$$

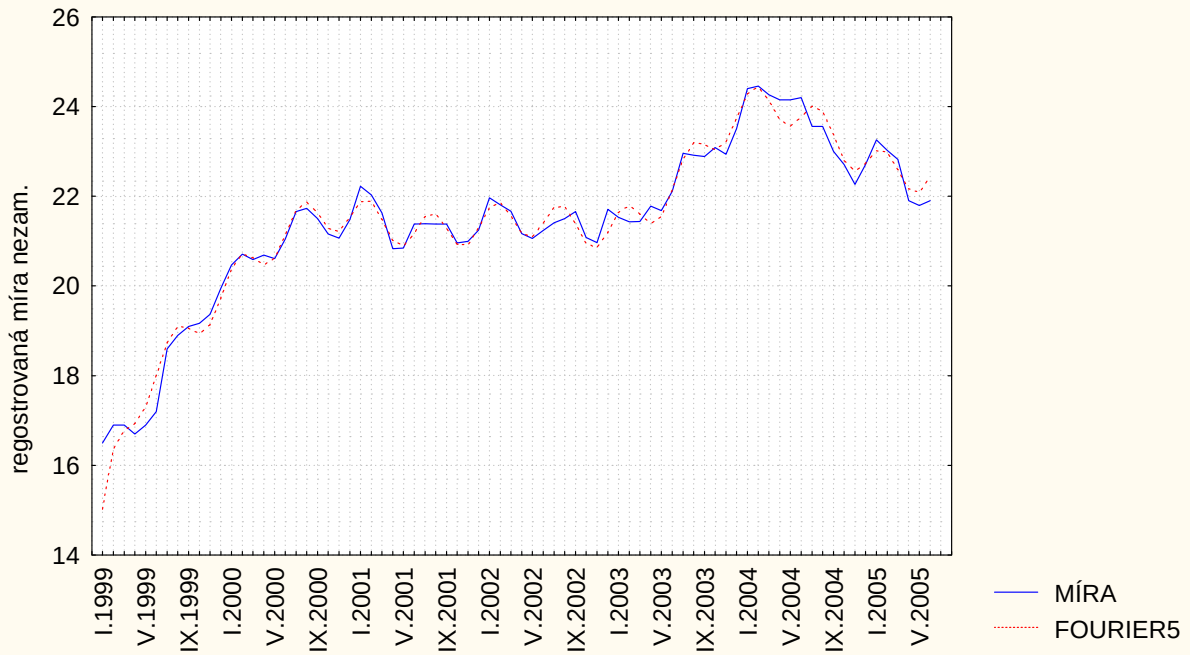
kde $H = \frac{n}{2}$ pro n sudé nebo $H = \frac{n-1}{2}$ pro n liché, $\omega_j = \frac{2\pi j}{n}$ pro $j = 1,....,H$ a ε_j jsou náhodné složky řady.

Od obou našich uvažovaných časových řad jsme tedy odečetli hodnoty příslušné trendové funkce a tím jsme dostali časové řady s konstantním trendem. Pomocí periodogramu a Fisherova testu jsme určili významné periody. V případě časové řady míry nezaměstnanosti mosteckého regionu vyšly významné periody 78;39;26;19,5 a 6 měsíců . Pomocí těchto významných period byly spočteny koeficienty a_j, b_j a po přičtení trendové funkce dostaneme aproximaci dané časové řady, která je zobrazena na grafu č.4. Příslušnou předpověď

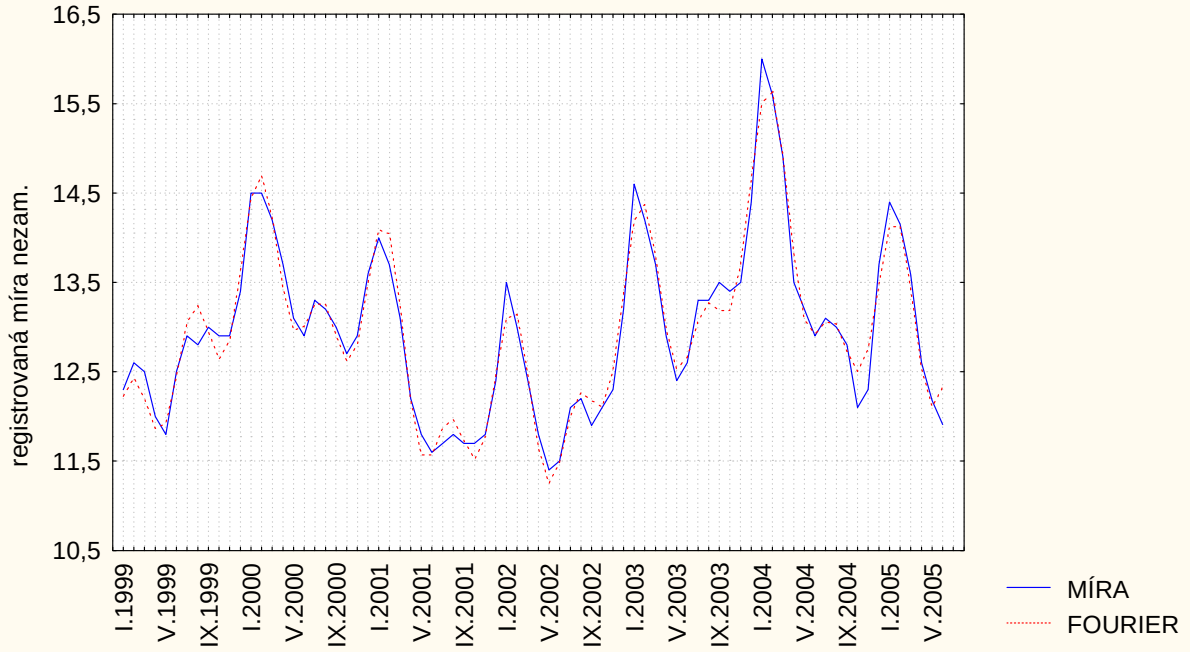
získáme dosazením hodnot 79-90 za t do příslušného trigonometrického polynomu. Hodnoty předpovědí pomocí této metody jsou uvedeny v druhém sloupci tabulky č.2.

Pro časovou řadu míry nezaměstnanosti regionu Litoměřice vyšly významné periody 78;39;26;15,6;13;11,14;9,75 a 6 měsíců. Aproximace dané časové řady je zobrazena na grafu č 5. Příslušná předpověď této časové řady je taky uvedena ve druhém sloupci tabulky č.2.

Graf 4.
Aproximace miry nezam.v regionu Most
pomocí trigonometrického polynomu



Graf 5
Aproximace míry nezam. v regionu LTM
pomocí trigonometrického polynomu



Zdroj: výpočty autora

Modely ARIMA,SARIMA

Box-Jenkinsova metodologie [1] je založena na myšlence, že hodnoty časové řady y_t jsou ovlivněny předchozími hodnotami této časové řady a náhodnými složkami. Předpokladem použití této metody je, aby časová řada byla stacionární. Tuto podmínku ani jedna z našich časových řad nesplňovala. Graf autoregresní funkce u obou časových řad totiž pomalu klesal a parciální autoregresní funkce měla významnou pouze první hodnotu. V obou případech jsme tedy naše časové řady stacionarizovali prvními diferencemi. Autoregresní funkce obou časových řad po transformaci první diferencí ukazovaly sezónní kolísání a v obou případech byl odladěn model SARIMA (1,1,0)(1,0,0). Předpovědi obou časových řad pomocí tohoto modelu jsou uvedeny v posledním sloupci tabulky č.2.

Tento způsob modelování jsme použili ještě jednou, ale na časové řady očištěné od sezónních odchylek. Takto očištěné časové řady opět vykazovaly nestacionaritu a bylo nutné je transformovat prvními diferencemi. Odladěný model je v tomto případě pro obě časové řady model ARIMA(1,1,0). K předpovědím získaným pomocí tohoto modelu jsme zase přičetli sezónní odchylky a získali tak předpovědi pro oba regiony, které jsou uvedeny ve čtvrtém sloupci tabulky č.2.

Prognózy

Předpovědi získané z výše popsaných metod jsou vedeny v následující tabulce č.2.

Tabulka č.2

Most				
měsíce	**FOURIER**	**TREND+SEZ**	**ARIMA+SEZ**	**SARIMA**
VII.05	22,85753602	23,3308939	22,30262731	21,69411552
VIII.05	22,96541280	23,3323350	22,18610782	21,69824193
IX.05	22,68543112	23,1943293	21,94557713	21,50287418
X.05	22,34560962	22,9124392	21,68735145	21,40143007
XI.05	22,34156611	22,7723658	21,57964225	21,24724781
XII.05	22,73978828	23,2255043	21,85846757	21,40148658
I.06	23,21107878	23,7093054	22,04763272	21,59427444
II.06	23,36052087	23,6201083	22,02652133	21,51015036
III.06	23,12365705	23,4015021	21,86624350	21,44004664
IV.06	22,83227500	23,1639924	21,56235460	21,11756901
V.06	22,88356012	23,0765844	21,40054904	21,08251710
VI.06	23,34304109	23,3754501	21,83221571	21,11756902
LTM				
měsíce	**FOURIER**	**TREND+SEZ**	**ARIMA+SEZ**	**SARIMA**
VII.05	12,82294096	13,0148893	12,35080753	12,07830802
VIII.05	13,03110109	12,9805264	12,30190275	11,99288653
IX.05	12,79203665	12,9031774	12,21129446	11,82255081
X.05	12,46311870	12,7324257	12,01045156	11,22650103
XI.05	12,52302665	12,8666044	12,14010882	11,39679881
XII.05	13,05707484	13,7090471	12,99909920	12,58888809
I.06	13,65622391	14,7803787	14,04720717	13,18493276
II.06	13,83473531	14,4646963	13,72701378	12,97205966
III.06	13,54145117	13,9323473	13,18615741	12,49522391
IV.06	13,23807718	13,1083317	12,34771117	11,65224642
V.06	13,44942253	12,6876493	11,91973466	11,29461961
VI.06	14,22435105	12,6178003	11,83399213	11,06471666

Zdroj: výpočty autora

Pro měsíce červenec a srpen jsou již známé skutečné hodnoty míry nezaměstnanosti. V mosteckém regionu tyto hodnoty jsou pro červenec 22,14 % a pro srpen 22,22%. Pro tuto časovou řadu je tedy nejpřesnější

metoda ARIMA(1,1,0), která byla provedena na časovou řadu očištěnou od sezónních odchylek a k prognóze pomocí této metody byly sezónní odchylky opět přičteny. V litoměřickém regionu jsou skutečné hodnoty míry nezaměstnanosti pro červenec 12,00% a pro srpen 12,02%. Tyto skutečné hodnoty jsou tedy nejblíže předpovědnímu modelu SARIMA(1,1,0)(1,0,0).

Literatura

[1] Arlt,J., Arltová,M., Rublíková,E.: *Analýza ekonomických časových řad s příklady*, 1,vyd., Praha: Fakulta informatiky a statistiky VŠE Praha, 2002, ISBN 80-245-0307-7
[2] Stuchlý,J.: *Statistické metody pro manažerské rozhodování.* 1.vyd., VŠE Praha, Fakulta managementu, nakladatelství Oeconomica 2004, ISBN 80-245-0153-8

Autor:

RNDr. Jana Šimsová
Katedra matematiky a statistiky
FSE UJEP Ústí nad Labem

ANALÝZA ROZPTYLU A SYSTÉM SAS®

Iveta Stankovičová

Súhrn
Analýza rozptylu (ANOVA) je štatistická metóda vhodná na vyhodnocovanie experimentálnych údajov. Procedúry ANOVA a GLM sú len dve procedúry, ktoré je možné použiť na analýzu rozptylu v systéme SAS/STAT®. Procedúra ANOVA je určená na analýzu vyvážených údajov, kým procedúra GLM je vhodná pre vyvážené aj nevyvážené dáta. Pre špeciálne prípady analýzy rozptylu sú určené procedúry MIXED, NESTED a NPAR1WAY.
Kľúčové slová
Analýza rozptylu (ANOVA), všeobecný lineárny model (GLM), efekty, viacnásobné porovnávanie, systém SAS®.
Summary
The analysis of variance (ANOVA) is statistical method for analyzing data from a wide variety of experimental designs. The ANOVA and GLM procedures are too of several procedures available in SAS/STAT® software for analysis of variance. The ANOVA procedure is designed to handle balanced data, whereas the GLM procedure can analyze both balanced and unbalanced data. The following procedures can be used for special analysis of variance: MIXED, NESTED and NPAR1WAY.
Key words
Analysis of variance (ANOVA), General Linear Model (GLM), Effects, Multiple Comparisons, SAS® System.

Úvod

Analýza rozptylu je štatistická metóda vhodná na skúmanie vzťahov medzi premennými a používa sa hlavne na vyhodnocovanie experimentálnych údajov. ANOVA je špeciálnym prípadom viacnásobnej regresie a teda aj špeciálnym prípadom všeobecného lineárneho modelu (GLM). Výpočtovo môže byť ANOVA interpretovaná ako regresia s umelými (dummy) vysvetľujúcimi premennými. Maticové riešenie umožňuje odvodiť pomerne jednoduché vzorce pre výpočet parametrov a testovanie hypotéz.
Vysvetľovanou (závislou) premennou v ANOVe je spojitá kvantitatívna premenná Y a ako faktor (vysvetľujúca, nezávislá premenná) vystupuje kategoriálna premenná A s malým počtom obmien. Podľa týchto obmien je možné hodnoty vysvetľovanej premennej triediť do skupín a tak skúmať, či faktorová premenná vplýva na vysvetľovanú premennú.
Analýzu rozptylu klasifikujeme z viacerých hľadísk. Ak skúmame vplyv len jedného faktora na vysvetľovanú premennú, tak ide o jednofaktorovú ANOVu. Jedná sa teda len o jednoduché triedenie. Pri viacerých faktoroch hovoríme o viacfaktorovej ANOVe (dvojité, trojité, ... triedenie). Podľa počtu vysvetľovaných premenných rozoznávame jednorozmernú (ANOVA) a viacrozmernú (MANOVA) analýzu rozptylu. ANOVu považujeme za vyváženú (balanced), ak v skupinách podľa hodnôt faktorovej premennej je rovnaký počet štatistických jednotiek. V opačnom prípade ide o nevyvážený (unbalanced) ANOVA model. Podľa toho, či analýzu rozptylu počítame z nameraných hodnôt závislej premennej alebo len z ich poradových (rank) hodnôt, rozoznávame parametrickú a neparametrickú ANOVu. Neparametrická ANOVA je vhodná hlavne pre malé súbory, v ktorých sa nedá overiť podmienka o normalite rozdelenia dát závislej premennej v podsúboroch podľa hodnôt faktorovej premennej.
V parametrických modeloch ANOVA uvažujeme, že hodnoty vysvetľovanej premennej Y je možné vyjadriť ako súčet neznámej strednej hodnoty μ, hodnoty A_i, ktorá vyjadruje tzv. efekt i-tej úrovne faktora a neznámej zložky ε, ktorá obsahuje ostatné vplyvy:

$$Y = \mu + A_i + \varepsilon \qquad i = 1,2, ..., k \tag{1}$$

Tento model je špeciálnym prípadom všeobecného lineárneho modelu (GLM) a môžeme ho vyjadriť v maticovom tvare:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2}$$

kde, **y** je vektor pozorovaní členený na *k* podvektorov, ktoré zodpovedajú úrovniam faktorovej premennej. Matica **X=[1, $\mathbf{X_A}$]** je tvorená stĺpcom jednotiek, ktoré patria k parametru μ a ďalších *k* stĺpcov, ktoré patria k parametrom A_1 až A_k, vektor **β** je vektor neznámych parametrov a **ε** je vektor neznámych zložiek.

Ak chceme odhadnúť parametre tohto modelu (2), čiže efekty jednotlivých úrovní faktora, tak sa nedá priamo použiť metóda najmenších štvorcov. Matica **X** nemá plnú hodnosť, t.j. h(X)<k+1 a preto je matica $\mathbf{X^TX}$ singulárna. Uvažuje sa preto naviac lineárne obmedzenie parametrov

$$\sum_{i=1}^{k} n_i A_i = 0 \tag{3}$$

a redukuje sa počet odhadovaných parametrov, alebo sa namiesto matice **X** typu *n* x *(k+1)* uvažuje s maticou kontrastov **X=[$\mathbf{I_n}$ $\mathbf{X_A C}$]** typu *n* x *k*, ktorá má už plnú hodnosť a tak získame rôzne typy kontrastov. Pre prvky matice **C**, t.j. pre koeficienty kontrastov platí

$$\sum_{i=1}^{k} c_{il} = 0, \qquad \sum_{i=1}^{k} {c_{il}}^2 > 0 \quad \text{kde } l = 1,2,...,k-1 \tag{4}$$

Namiesto parametrov A_i potom uvažujeme parametre A_i^*, ktoré sú navzájom vo vzťahu

$$A_i = \sum_{l=1}^{k-1} c_{il} A_l^* = 0, \qquad \text{kde } i = 1,2,...,k \tag{5}$$

Voľba typu kontrastov má vplyv na interpretáciu parametrov modelu, ale nie na výsledky testovania zložených hypotéz o parametroch modelu.

Ak preukážeme existenciu vplyvu faktora, tak môže nasledovať hlbšia analýza výsledkov, v ktorej zisťujeme, medzi ktorými skupinami existujú rozdiely. Okrem hypotézy H_0: $\mu_i - \mu_j = 0$ môžeme testovať aj hypotézu o lineárnej kombinácii stredných hodnôt, t.j. o nulovom lineárnom kontraste.

$$\Psi_i = \sum_{i=1}^{k} c_i \mu \qquad \text{kde} \quad \sum_{i=1}^{k} c_i = 0 \tag{6}$$

Pri testovaní kontrastov je podstatné, či bol kontrast zvolený dopredu alebo až na základe výsledkov. Pri simultánnom teste viacerých kontrastov aj to, či sú kontrasty nezávislé. Hrozí zvýšenie chyby I. druhu, to znamená, že vo viacej ako 100α% prípadov môže byť zamietnutie testovanej hypotézy dôsledkom náhodného kolísania a nie existujúceho rozdielu medzi strednými hodnotami. V teórii bolo odvodených veľa metód, ktoré kontrolujú chybu I. druhu (α) a ktoré sa označujú ako metódy viacnásobného porovnávania (multiple comparisons). Podstatou všetkých je konštrukcia 100(1-α)% intervalu spoľahlivosti pre kontrast Ψ. Ak tento interval neobsahuje 0, preukázali sme nenulovosť kontrastu na hladine významnosti α.

Jednotlivé metódy sa líšia jednak podmienkami použitia, jednak šírkou intervalu spoľahlivosti a sú súčasťou štatistických počítačových programov ako je aj systém SAS®.

Analýza rozptylu v systéme SAS®

Systém SAS® umožňuje vykonávať analýzu rozptylu pomocou niekoľkých procedúr modulu SAS/STAT. Sú to procedúry:

- ANOVA - len pre vyvážené modely,
- GLM - pre vyvážené ale aj nevyvážené modely,
- MIXED – pre zmiešané modely,
- NESTED – pre hierarchické modely,
- NPAR1WAY - pre neparametrickú ANOVu.

Procedúra ANOVA je určená len pre vyvážené modely so špeciálnou štruktúrou a preto je rýchlejšia a vyžaduje menej pamäte ako procedúra GLM. Nie je však vhodná na jednorozmernú analýzu rozptylu experimentov v tvare latinských štvorcov, na čiastočne vyvážené neúplné blokové experimenty, na úplné hierarchické (nested) experimenty a na experimenty s takými početnosťami v skupinách, ktoré sú proporcionálne navzájom a tiež vo vzťahu k základnému súboru. Tieto výnimky majú taký dizajn, že všetky faktory sú navzájom ortogonálne. PROC ANOVA je vhodná pre experimenty, ktorých blokové diagonálne matice $\mathbf{X}^T\mathbf{X}$ obsahujú prvky vo všetkých blokoch s rovnakými hodnotami a parciálne testuje túto požiadavku rovnosti stredných hodnôt v blokoch. Bohužiaľ, pre niektoré typy experimentov je tento test v PROC ANOVA chybný. Ak test zamietne hypotézu o rovnosti stredných hodnôt, tak procedúra vypíše varovanie, že analýza nie je platná a experiment nie je vyvážený a mali by sme radšej použiť procedúru GLM. Pre úplné overenie tejto podmienky je potrebná úplná matica $\mathbf{X}^T\mathbf{X}$**,** a preto ak nie ste si istý splnením tejto podmienky, tak by ste mali použiť PROC GLM.
Na nasledujúcom príklade demonštrujeme, že výsledky F štatistiky pomocou procedúr ANOVA a GLM môžu byť rôzne, keď dáta predstavujú nevyvážený model.

SAS kód na vytvorenie dátového súboru WORK.EXP:

```
data exp;
        input A $ B $ Y @@;
        datalines;
                A1 B1 12 A1 B1 14     A1 B2 11 A1 B2 9
                A2 B1 20 A2 B1 18     A2 B2 17 ;
```

Dátový súbor WORK.EXP:

A	B	Y
A1	B1	12
A1	B1	14
A1	B2	11
A1	B2	9
A2	B1	20
A2	B1	18
A2	B2	17

SAS kód pre procedúru ANOVA:

```
PROC ANOVA DATA=WORK.EXP;
        CLASS A B;
        MODEL Y = A B;
RUN;
```

SAS výstup pre procedúru ANOVA (časť):

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	97,71	48,86	Infty	<,0001
Error	4	0,00	0,00		
Corrected Total	6	97,71			

Výstraha v LOG okne pre nevyvážený model:

```
WARNING: PROC ANOVA has determined that the number of
observations in each cell is not equal. PROC GLM may be more
appropriate.
```

Source	DF	Anova SS	Mean Square	F Value	Pr > F
A	1	80,05	80,05	Infty	<,0001
B	1	23,05	23,05	Infty	<,0001

SAS kód pre procedúru GLM:

```
PROC GLM DATA=WORK.EXP;
        CLASS A B;
        MODEL Y = A B;
RUN;
```

SAS výstup pre procedúru GLM (časť):

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	91,31	45,66	28,54	0,0043
Error	4	6,40	1,60		
Corrected Total	6	97,71			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	1	80,05	80,05	50,03	0,0021
B	1	11,27	11,27	7,04	0,0568

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	1	68,27	68,27	42,67	0,0028
B	1	11,27	11,27	7,04	0,0568

Určenie efektov v analýze rozptylu v systéme SAS®

V modeloch analýzy rozptylu môžeme určiť efekty, ktoré sú kombináciou faktorových premenných použitých na vysvetlenie variability závislej premennej v nasledovných formách:

- ako hlavné (main) efekty:

```
proc glm data=WORK.EXP;
   class A B;
```

Dáta	Dizajn matice X		
		A	B

```
   model Y=A B;
run;
```

A	B	μ	A1	A2	B1	B2
1	1	1	1	0	1	0
1	2	1	1	0	0	1
2	1	1	0	1	1	0
2	2	1	0	1	0	1

- ako krížové (interaction) efekty:

```
proc glm data=WORK.EXP;
   class A B;
   model Y=A B A*B;
run;
```

Dáta		Dizajn matice X								
			A		B		A*B			
A	B	μ	A1	A2	B1	B2	A1B1	A1B2	A2B1	A2B2
1	1	1	1	0	1	0	1	0	0	0
1	2	1	1	0	0	1	0	1	0	0
2	1	1	0	1	1	0	0	0	1	0
2	2	1	0	1	0	1	0	0	0	1

- ako hierarchické (nested) efekty:

```
proc glm data=WORK.EXP;
   class A B;
   model y=A B(A);
run;
```

Dáta		Dizajn matice X						
			A		B(A)			
A	B	μ	A1	A2	B1A1	B2A1	B1A2	B2A2
1	1	1	1	0	1	0	0	0
1	2	1	1	0	0	1	0	0
2	1	1	0	1	0	0	1	0
2	2	1	0	1	0	0	0	1

Procedúry GLM a ANOVA umožňujú vypočítať všetky stredné hodnoty závislej premennej pre efekty, ktoré sa nachádzajú na pravej strane príkazu MODEL a to rôznymi metódami zadanými ako voľba v príkaze MEANS (viď tab. 1 a 2). Je potrebné upozorniť, že procedúra ANOVA netestuje viacnásobné porovnania pre krížové efekty. Na tento účel je potrebné použiť príkaz LSMEANS v procedúre GLM.

Viacnásobné porovnávania

V analýze rozptylu F štatistika v ANOVA tabuľke testuje globálnu hypotézu H_0: $\mu_1=\mu_2=...=\mu_k$ Netestuje však, ktoré stredné hodnoty sú navzájom rozdielne. K tomuto účelu sa používajú procedúry viacnásobného porovnávania (multiple comparisons), ktoré klasifikujeme dvoma spôsobmi:

1. podľa toho ako sa vykonávajú:
   - porovnávanie všetkých párov stredných hodnôt, t.j. aritmetických priemerov,
   - porovnávanie medzi kontrolným a všetkými ostatnými strednými hodnotami,
2. podľa toho aký silný úsudok poskytujú[1]:
   - individuálny: rozdiely medzi strednými hodnotami neupravené o multiplicity,
   - nehomogénnosť: stredné hodnoty sú rozdielne,
   - nerovnosť: ktoré stredné hodnoty sú rozdielne,
   - intervaly: paralelné intervaly spoľahlivosti pre rozdiely stredných hodnôt.

Sila úsudku nám hovorí, ako je možné uvažovať o štruktúre stredných hodnôt, keď F test je významný. Súvisí to s kontrolou chyby I. druhu (α) pri viacnásobnom porovnávaní. Metódy, ktoré kontrolujú len individuálne chyby I. druhu (α), t.j. kontrolujú chybu spôsobu porovnávania (CER = comparisonwise error rate), nie sú skutočné metódy viacnásobného porovnávania.
Problém nastáva pri opakovaných t-testoch. Ak predpokladáme, že máme k=10 stredných hodnôt a každý t-test vykonávame na hladine významnosti α=0,05, potom pri viacnásobnom porovnávaní máme až $n_c=10(10-1)/2=45$ párov na porovnanie a každý s pravdepodobnosťou 0,05, že sa dopustíme chyby I. druhu (t.j. že nesprávne zamietneme nulovú hypotézu o zhode stredných hodnôt). Šanca, že najmenej raz sa pri týchto 45-tich pároch dopustíme chyby I. druhu je omnoho vyššia ako 0,05. Je ťažké presne vypočítať túto chybu I. druhu pri viacnásobných porovnávaniach, ale môžeme vypočítať aspoň pesimistický odhad tejto chyby za predpokladu, že porovnávania sú

[1] Úsudky sú zoradené od najslabšieho po najsilnejší.

nezávislé. Vypočítame hornú hranicu chyby, ktorú nazývame chybou spôsobu experimentu (EER = experimentwise error rate) podľa vzťahu $1-(1-\alpha)^{n_c}$, čiže v našom prípade pre 10 stredných hodnôt je to až $1-(1-0{,}05)^{45}=0{,}90$. Pri zvyšovaní počtu stredných hodnôt sa táto chyba blíži k číslu 1. V teórii štatistiky bolo odvodených veľa metód, ktoré môžu udržať túto chybu na nízkej hodnote. V nasledovných tabuľkách uvádzame prehľad metód umožňujúcich viacnásobné porovnávania v procedúre GLM v príkazoch MEANS a LSMEANS podľa sily úsudku.

Tab. 1: Porovnávanie všetkých párov stredných hodnôt

Metóda	Sila úsudku	Syntax	
		MEANS	LSMEANS
Student's *t*	individuálny	T	PDIFF ADJUST=T
Duncan	individuálny	DUNCAN	
Student-Newman-Keuls	nehomogénnosť	SNK	
REGWQ	nerovnosť	REGWQ	
Tukey-Kramer	intervaly	TUKEY	PDIFF ADJUST=TUKEY
Bonferroni	intervaly	BON	PDIFF ADJUST=BON
Sidak	intervaly	SIDAK	PDIFF ADJUST=SIDAK
Scheffé	intervaly	SCHEFFE	PDIFF ADJUST=SCHEFFE
SMM	intervaly	SMM	PDIFF ADJUST=SMM
Gabriel	intervaly	GABRIEL	
Simulation	intervaly		PDIFF ADJUST=SIMULATE

Tab. 2: Porovnávanie medzi kontrolným a všetkými ostatnými strednými hodnotami

Metóda	Sila úsudku	Syntax	
		MEANS	LSMEANS
Student's *t*	individuálny		PDIFF=CONTROL ADJUST=T
Dunnett	intervaly	DUNNETT	PDIFF=CONTROL ADJUST=DUNNETT
Bonferroni	intervaly		PDIFF=CONTROL ADJUST=BON
Sidak	intervaly		PDIFF=CONTROL ADJUST=SIDAK
Scheffé	intervaly		PDIFF=CONTROL ADJUST=SCHEFFE
SMM	intervaly		PDIFF=CONTROL ADJUST=SMM
Simulation	intervaly		PDIFF=CONTROL ADJUST=SIMULATE

Záver

Dnes si nevieme predstaviť, že by sme robili štatistické analýzy bez vhodného softvéru. Na druhej strane, treba užívateľov vystríhať od používania procedúr bez dôkladného preštudovania si dokumentácie. Často sa stáva, že užívatelia nepoužívajú vhodné procedúry, alebo nesprávne interpretujú výsledky počítačových výstupov.

Použitá literatúra

[1] Bohdalová M., Wimmer G., Witkovský V., Savin A: Interval spoľahlivosti pre referenčnú hodnotu v medzilaboratórnych porovnávacích štúdiách. Zborník konferencie SAS Forum 2004, Bratislava 17.5.2004, str.1-4.

[2] Hebák Petr, Hustopecký Jiří, Jarošová Eva, Pecáková Iva: Vícerozměrné statistické metody 1. INFORMATORIUM. Praha 2004. ISBN 80-7333-025-3.

[3] Keith E. Muller, Bethel A. Fetterman: Regression and ANOVA. An Integrated Approach Using SAS® Software. Copyright © 2002 by SAS Institute Inc., Cary, NC, USA, ISBN 1-58025-890-5.

[4] Luha Ján a kol.: Niektoré možnosti využitia analýzy rozptylu pri skúmaní ekozložky životného prostredia. Príspevok na seminári "Ekologické problémy urbanizovaného prostredia". PFUK Bratislava 1988.

[5] SAS OnLine Documentation: http://support.sas.com/91doc/docMainpage.jsp

Kontaktná adresa: Ing. Iveta Stankovičová, PhD.
Katedra informačných systémov
Fakulta managementu UK
Odbojárov 10, 820 05 Bratislava e-mail: iveta.stankovicova@fm.uniba.sk

UKÁŽKA VYUŽITIA SAS-u PRI HODNOTENÍ GENETICKÝCH ZDROJOV

Beáta Stehlíková

Abstract
The aim of the paper is to describe SAS as useful tool for genetic resources analysis. Simple statistical methods – histogram and kernel density are exploits to detection of the two various groups.

Key words: SAS, kernel density, histogram, genetic resources

Postavenie štatistiky excelentne vyjadril T. Miština na XII. letnej škole biometriky, keď píše, že výskumný pracovník v poľnohospodársve skúma živé organizmy, ich chovanie v prostredí a hľadá prostriedky na odlíšenie náhodných vplyvov od vplyvu faktorov, ktoré reguluje pestovateľ, chovateľ, alebo od vplyvu faktorov, ktoré sú na poľnohospodárovi nezávislé. Zdá sa, že doteraz sa nenašiel lepší nástroj na toto rozlišovanie, ako je matematika hodnotiaca náhodné javy - štatistika používaná v biológii. SAS je prostriedkom umožňujúcim intenzívne využitie štatistických metód pri hodnotení biologického materiálu.

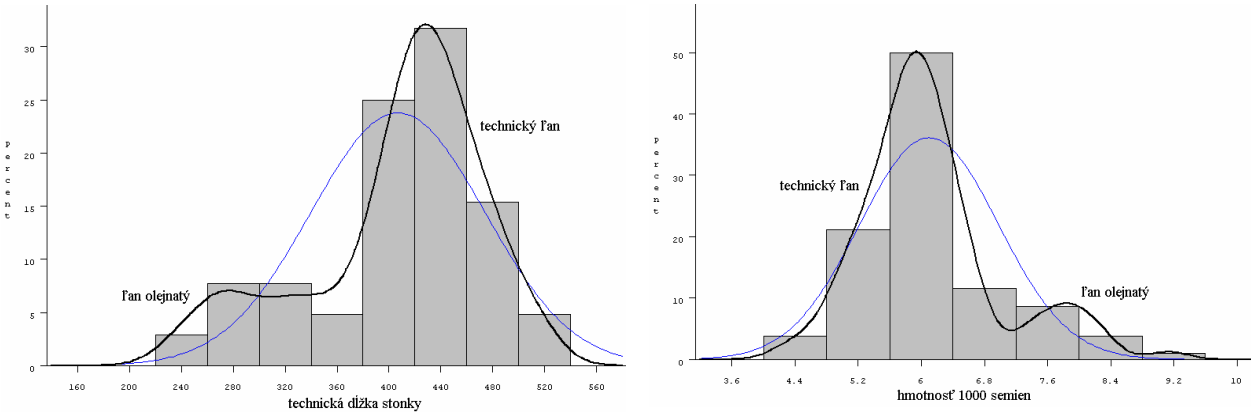
Materiál a metódy

Hodnotený súbor obsahuje údaje o technickej dĺžke stonky (Tdston) a hmotnosti 1000 semien (Vsem) 24 genotypov olejnatých a 80 genotypov technického ľanu. Údaje poskytol Inštitút ochrany biodiversity a biologickej bezpečnosti Slovenskej poľnohospodárskej university v Nitre. Cieľom príspevku je ukázať ako s využitím SAS-u efektívnejšie hodnotiť biologický materiál. V rámci popisnej štatistiky sa obvykle konštruuje histogram pre každý z hodnotených znakov a testuje sa zhoda empirického rozdelenia s teoretickým, obvykle normálnym. Pre lepší popis hodnotených údajov okrem histogramu sa doporučuje skonštruovať aj jadrovú funkciu hustoty. Nech ξ je náhodná premenná s funkciou hustoty $f(x)$. Na základe výberového súboru je potrebné skonštruovať funkciu $\tilde{f}_n(x)$, tzv. jadrovú funkciu, ktorá je odhadom $f(x)$.

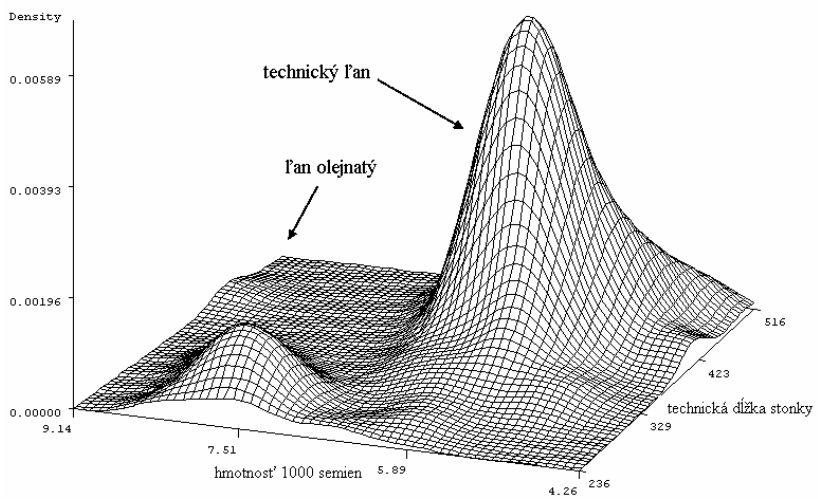
Výsledok a diskusia

Pri hodnotení biologického materiálu sa obvykle konštruuje histogram s následným testovaním zhody empirického rozdelenia s niektorým teoretickým rozdelením. Technická dĺžka stonky ani hmotnosť 1000 semien 104 genotypov ľanu nemá normálne rozdelenie. Skupinu 104 genotypov tvorí 24 genotypov ľanu olejnatého a 80 genotypov technického ľanu. V prípade genotypov pochádzajúcich zo zberových expedícií podobná informácia chýba. Bez znázornenia jadrovej funkcie by pravdepodobne zostalo pri konštatovaní nenormality hodnotených javov. Znázornenie jadrovej funkcie však v prípade oboch hodnotených znakov indikuje existenciu dvoch skupín genotypov. Grafické znázornenie hodnotených genotypov pomocou trojrozmerného grafu umožňuje identifikovať skupinu genotypov technického ľanu s vyššou technickou dĺžkou stonky a nižšou hmotnosťou 1000 semien. Pre olejnaté genotypy ľanu je typická vyššia hmotnosť 1000 semien, nakoľko sa zo semien lisuje ľanový olej.
SAS umožňuje konštrukciu histogramu a jadrovej funkcie pomocou procedúry *univariate*. Výpis programu je na konci príspevku.

Obrázok 1 Histogram a jadrová funkcia pre technickú dĺžku stonky a hmotnosti 1000 semien



Obrázok 2 Závislosť technickej dĺžky stonky a hmotnosti 1000 semien



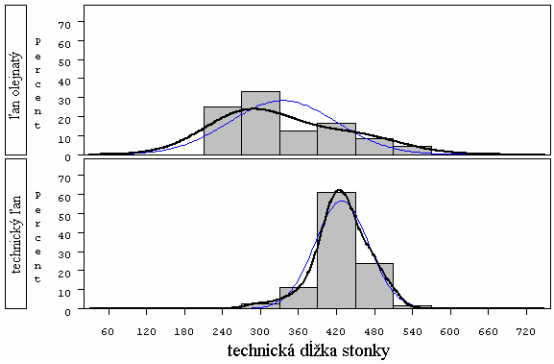
Výpis programu v SASe

```
proc univariate data=lan normaltest NOPRINT;
var tdston;
HISTOGRAM / kernel(w=2 color=black) cfill=ltgray NORMAL ;
run;
proc univariate data=lan normaltest NOPRINT;
var vsem;
HISTOGRAM / kernel(w=2 color=black) cfill=ltgray NORMAL ;
run;
PROC SORT data=lan;
by hv;
run;
proc univariate data=lan normaltest NOPRINT;
var tdston;
class hv;
HISTOGRAM / kernel(w=2 color=black) cfill=ltgray NORMAL ;
run;
proc univariate data=lan normaltest NOPRINT;
var vsem;
class hv;
HISTOGRAM / kernel(w=2 color=black) cfill=ltgray NORMAL ;
run;
proc kde data=lan out=vyslan bwm=0.7,0.7 ;
var vsem tdston ;
```

```
run;
proc g3d data=vyslan;
plot vsem * tdston=density;
run;
```

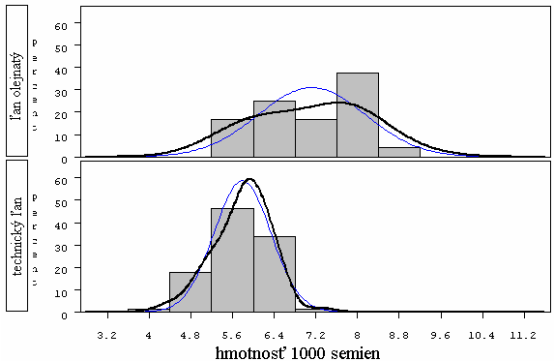
Pre úplnosť dodajme, že v oboch podsúboroch, ktoré vzniknú rozdelením hodnotených genotypov podľa hospodárskeho využitia majú oba hodnotené znaky rozdelenie, ktoré sa štatisticky preukazne nelíši od normálneho.

Obrázok 3 Test normality technickej dĺžky stonky v podsúboroch



Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.17027798	Pr > D	0.072
Kolmogorov-Smirnov	D	0.08898621	Pr > D	0.118

Obrázok 4 Test normality hmotnosti 1000 semien v podsúboroch



Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.15926707	Pr > D	0.114
Kolmogorov-Smirnov	D	0.08867957	Pr > D	0.121

Záver

Na hodnotenom súbore bolo sledovaných ďalších štrnásť znakov. Diskriminačná analýza pomocou analyzovaných dvoch znakov je rovnako dobrá ako diskriminácia podľa všetkých šestnástich znakov (celková pravdepodobnosť nesprávneho zaradenia je v oboch prípadoch zhodne 15,8 percent). Pozorná analýza pravdepodobnostného rozdelenia znakov umožňuje tiež identifikáciu znakov vhodných pre diskrimináciu.

Príspevok je súčasťou riešenia projektu KEGA 3/2060/04.

Literatúra

Morrison, D.F.: *Multivariate statistical methods*. McGraw-Hill, 1981. ISBN 0-07-085469-6

Kontaktná adresa

doc. RNDr. Beáta Stehlíková, CSc., Katedra štatistiky a operačného výskumu, FEM SPU v Nitre, Tr. A. Hlinku 2, 949 76 Nitra. e-mail: Beata.Stehlikova@uniag.sk

VYUŽITIE ATRIBÚTU FARBY PRI HODNOTENÍ GENETICKÝCH ZDROJOV

Beáta Stehlíková

Abstract
Colour is one attribute of the genetic resources. The aim of the paper is to characterize colour as the triangular fuzzy number. This approach allows operate with colour as quantitative variable in the genetic resources database.

Key words: colour, fuzzy number, genetic resources

Dvadsiatim druhým cieľom Národnej stratégie ochrany biodiverzity na Slovensku je ustanovenie celoštátneho mechanizmu "clearing-house" vzťahujúceho sa na biodiverzitu. Medzi strategické smery patrí posilniť databázy o biodiverzite ako i úloha zaviesť využitie nových technológií v manažmente dát. Podľa Národnej stratégie ochrany biodiverzity existuje na Slovensku niekoľko databáz, ktoré sa týkajú biodiverzity. Genotypdata, informačný systém pre dokumentáciu a vyhodnotenie genetických zdrojov pestovaných druhov, pre výskum, pestovanie a komerčné využitie s obrazovou dokumentáciou, bol založený a pracuje na SPU v Nitre. Okrem obrazovej dokumentácie týkajúcej sa genetických zdrojov je potrebné pracovať s atribútom farba charakterizovaným pomocou čísla.
Farby sú naše vnemy. Ľubovoľný zdroj svetla sa dá analyzovať pomocou difrakčnej mriežky alebo hranolu a dá sa určiť jeho spektrálne zloženie, t.j. množstvo svetla každej vlnovej dĺžky. Fyzikálne sa dá všetko veľmi presne určiť. Inou otázkou je, akú farbu budeme vnímať. Je známe, že existuje viac spektrálnych rozložení, ktoré spôsobujú rovnaký vizuálny efekt. Na určenie farieb existuje viacero farebných atlasov. Najznámejší je Munsellov atlas (Munsell Color Charts for Plant Tissues), ďalej Methuen Handbook of Colour, Munsell Color Charts for Plant Tissues, farby RHS (Royal Horticultural Society), ktorý obsahuje farby zoradené podľa tónu farby, čistoty farby a jasu.

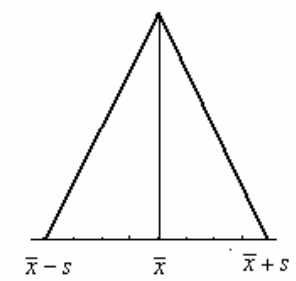
Materiál a metódy
Pre reprezentáciu farby sa najčastejšie používa histogram. Na histograme je každej úrovni jasu priradená zodpovedajúcu početnosť pixelov daného jasu v obraze. Formálne to znamená, že histogram obrazu W je definovaný ako vektor H(W) = $\left(h_{c_1}, h_{c_2}, ..., h_{c_n}\right)$, kde h_{c_i} , (i =1, 2, ..., n) je počet bodov s jasovou hodnotou c_i v obraze W, n je počet hladín a $n_W = \sum_{i=1}^{n} h_{c_i}$ je počet bodov obrazu W. Vzdialenosť medzi histogramom obrazu W a histogramom obrazu Q môžeme vypočítať v euklidovej metrike alebo v metrike l_1. Popri histograme môžu byť použité aj ďalšie štatistické reprezentácie farby. Matematickým základom myšlienky metódy momentov farieb je skutočnosť, že každé rozdelenie farieb môže byť charakterizované jej momentmi. Väčšina podstatnej informácie je sústredená v momentoch nižšieho rádu – v priemeri, rozptyle a šikmosti. Obraz W s histogramom H(W) sa z hľadiska farby môže reprezentovať ako trojuholníkové fuzzy číslo (a_1, a_2, a_3) = $(\bar{x} - s, \bar{x}, \bar{x} + s)$, kde $\bar{x} = \sum_{i=1}^{n} c_i h_{c_i} / n_W$ je priemerná hodnota, s je smerodajná odchýlka.
Rozdiel medzi dvoma farbami je možné definovať ako vzdialenosť dvoch fuzzy čísiel (a_1, a_2, a_3) a (b_1, b_2, b_3). Vzdialenosť môže byť definovaná pomocou euklidovskej metriky

$$d_{eukl} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2} .$$

Hagaman zaviedol vzdialenosť dvoch fuzzy čísiel spôsobom, v ktorom rozdiely hodnôt s funkciou príslušnosti blízko 1 majú väčšiu váhu ako rozdiely rozdiely hodnôt s funkciou príslušnosti blízko nuly. Hagamanova vzdialenosť[1] dvoch fuzzy čísiel A= (a_1, a_2, a_3) a B = (b_1, b_2, b_3)

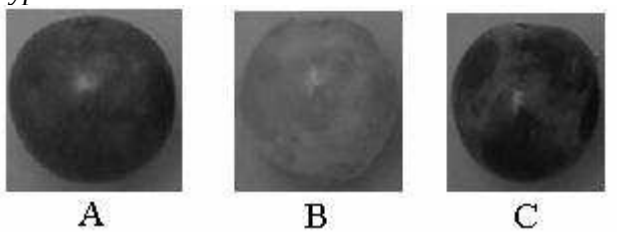
$$d^2_{hagaman} = (a_2 - b_2)^2 + \frac{1}{3}(a_2 - b_2)(a_3 - 2a_2 + a_1 - b_3 + 2b_2 - b_1) + \\ + \frac{1}{12}((a_2 - a_1 - b_2 + b_1)^2 + (a_3 - a_2 - b_3 + b_2)^2).$$

Obrázok 1 Trojuholníkové fuzzy číslo



Aplikáciu uvedeného postupu budeme demonštrovať na príklade bobúľ troch genotypov viniča A, B, C. Našou úlohou je rozhodnúť, ktoré farby bobúľ viniča sú najviac podobné.

Obrázok 2 Bobule troch genotypov viniča



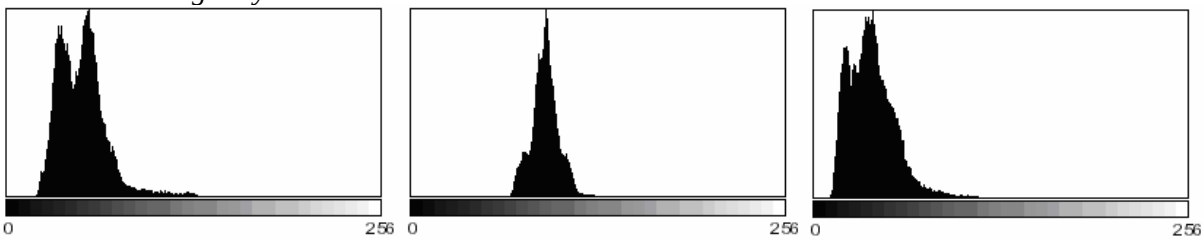
Pri výpočtoch bol použitý softvér Mathematica a ImageJ.

Výsledok a diskusia

Tabuľka 1 Charakteristiky bobúľ viniča z obrázku 2

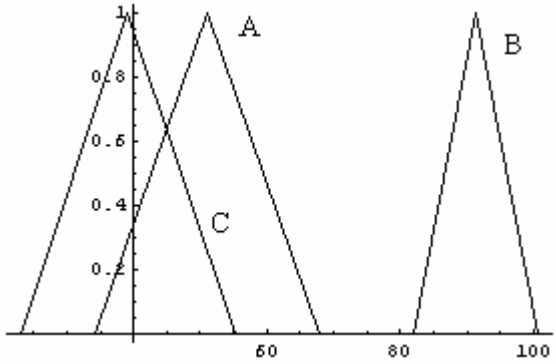
Charakteristika	A	B	C
priemer	51,057	91,353	39,084
smerodajná odchýlka	16,823	9,404	16,105
min	16	36	8
max	138	139	123
R	122	103	115
modus	56	92	40
podiel pixelov v moduse	2,966	1,903	2,908

Obrázok 3 Histogramy



[1] presnejšie povedané Hagamanova vzdialenosť d(A,B,f(q)) pre fuzzy čísla A a B pri voľbe funkcie f(q)=q. Podrobnejšie Bardosi, Duckstein, 1995

Obrázok 4 Trojuholníkové fuzzy čísla zodpovedajúce bobuliam troch genotypov A, B,C



Tabuľka 2 Euklidova vzdialenosť, Hagamanova vzdialenosť,vzdialenosť histogramov podľa metriky l_1.

	A	B	C
A		70,58	**20,76**
B	70,58		91,03
C	20,76	91,03	

A	B	C
	40,41	**11,98**
40,41		52,34
11,98	52,34	

A	B	C
	1,79	**0,59**
1,79		1,91
0,59	1,91	

V každej z použitých metód je z hľadiska farby najpodobnejšia bobuľa A a C a najmenej podobné sú bobule B a C.

Záver

Popis farby ako trojuholníkového fuzzy čísla umožňuje zaradiť informáciu o farbe do databázy obsahujúcej číselné údaje o genotypoch, čo umožňuje spracovanie informácie o farbe klasickými metódami prehľadávania databáz s cieľom identifikácie genetických zdrojov určitých vlastností.

Príspevok je súčasťou riešenia projektu KEGA 3/2060/04.

Literatúra

Kosko, B. : Fuzzy thinking: the new science of fuzzy logic. London: Flamingo, 1994
Siler, William: Constructing Fuzzy http://members.aol.com/wsiler/intro.htm (17.11.2005)

Kontaktná adresa

doc. RNDr. Beáta Stehlíková, CSc., Katedra štatistiky a operačného výskumu, FEM SPU v Nitre, Tr. A. Hlinku 2, 949 76 Nitra. e-mail: Beata.Stehlikova@uniag.sk

Využitie predajno-kúpnej parity pri stanovení spravodlivej ceny opcie na akciu

Vladimír Úradníček

Abstrakt
Contribution consider an important relationship between put and call option for a non-dividend-paying stock. This realtionship is known as put-call parity. It shows that the value of a European put with a certain exercise price and exercise date can be deduced from the value of a European call with the same exercise price and date, and vice versa. To illustrate, we consider a European put option on a non-dividend-paying stock of 2006 March SAP OH-E when the strike price is 40 USD.

1. Úvod

Cieľom príspevku je ilustrovať možnosť aplikácie predajno-kúpnej parity pre stanovenie spravodlivej ceny opcie na akciu SAP. Úlohy uvedeného typu môžu riešiť študenti študijného programu Financie, bankovníctvo a investovanie v rámci predmetu Finančná matematika s pomocou softvérového produktu MS Excel, pri používaní reálnych dát z internetových stránok chicagskej burzy (CBOE) a http://finance.yahoo.com.

2. Teoretické východiská

Majme portfólio, ktoré je tvorené jednou akciou (cena akcie je *S)*, jednou predajnou (*P*) a jednou kúpnou (*C*) opciou európskeho typu. Predpokladajme, že predajná aj kúpna opcia majú v čase vypršania (*T*) rovnakú realizačnú cenu (*K*).
V čase vypršania opcií platí:

$$S_T + P_T^K - C_T^K = K \qquad (1)$$

O platnosti vzťahu (1) sa môžeme ľahko presvedčiť. Platí, že hodnoty príslušných derivátov v čase vypršania sú: $P_T = (K - S_T)^+$ a $C_T = (S_T - K)^+$. Potom,
ak $K > S_T$, tak platí:

$$S_T + K - S_T - 0 = K,$$

ak $K < S_T$, tak platí:

$$S_T + 0 - S_T + K = K$$

a ak $K = S_T$, tak platí:

$$S_T + 0 - 0 = K$$
$$K = K$$

Vzťah (1) môžeme vo všeobecnosti zapísať pre čas *t*, $0 < t < T$:

$$S_t + P_t^K - C_t^K = K.e^{-r(T-t)}, \qquad (2)$$

resp. pre súčasnosť ($t = 0$) :

$$S_0 + P_0^K - C_0^K = K.e^{-rT}. \qquad (3)$$

Vzťahy (2), resp. (3) môžeme využiť pre stanovenie hodnoty predajnej opcie európskeho typu, ak poznáme hodnotu kúpnej opcie európskeho typu . Potom:

$$P_t^K = C_t^K + K.e^{-r(T-t)} - S_t, \text{ resp.}$$
$$P_0^K = C_0^K + K.e^{-rT} - S_0.$$

Vzťah (2) môžeme považovať za najvšeobecnejší zápis pre predajno-kúpnu paritu (put-call paritu) európskych predajno-kúpnych opcií. Treba zdôrazniť, že uvedené parity neplatia pre americké opcie (opcie amerického typu). Dá sa ukázať, že v ich prípade put-call parita je v tvare nerovností[1]:

$$K > S_t + P_t^{KA} - C_t^{KA} > K.e^{-r(T-t)}. \qquad (4)$$

3. Aplikácia predajno-kúpnej parity pre stanovenie hodnoty opcie na akciu európskeho typu

Pre aplikáciu predajno-kúpnej parity sme zvolili príklad stanovenia spravodlivej hodnoty budúcoročnej marcovej (2006) európskej predajnej opcie na akciu SAP neprinášajúcu dividendu s realizačnou cenou K = 45 USD. Predpokladajme, že máme k dispozícii spravodlivú hodnotu kúpnej opcie uvedených charakteristík, ktorú sme vypočítali pomocou Black-Scholesových vzorcov. Celý výpočet realizujeme 23. novembra 2005.
Ak označíme čas vypršania T, cenu akcie v čase vypršania S_T, dohodnutú cenu opcie K a bezrizikovú úrokovú mieru r, potom cenu európskej kúpnej opcie na akcie neprinášajúcej dividendy C^E môžeme definovať ako diskontovanú strednú hodnotu

$$C^E = e^{-rT}.E_Q\left[(S_T - K)^+\right] \qquad (5)$$

Dá sa ukázať, že Black-Scholesov vzorec pre hodnotu európskej kúpnej opcie na akciu neprinášajúcu dividendy v čase uzatvorenia kontraktu má potom tvar:

$$C^E = V(S,0) = S_0.\Phi\left(\frac{\ln\frac{S_0}{K} + rT + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\right) - K.e^{-rT}\Phi\left(\frac{\ln\frac{S_0}{K} + rT - \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\right) =$$
$$= S_0.\Phi(d_1) - K.e^{-rT}\Phi(d_2), \qquad (6)$$

kde $\Phi(d_1)$ je hodnota distribučnej funkcie N(0;1) rozdelenia v bode d_1 a

$\Phi(d_2)$ je hodnota distribučnej funkcie N(0;1) rozdelenia v bode d_2.

Za účelom výpočtu jednotlivých charakteristík využijeme informácie z webových stránok: http://finance.yahoo.com , resp. http://bonds.yahoo.com . Parameter σ^2 – ročnú volatilitu ceny akcie sme odhadli z historických cien za obdobie od 22.8.2005 do 22.11.2005. Nevychýlený odhad rozptylu relatívnych prírastkov ceny akcie mal v prepočte na rok hodnotu *0,18414*[2]. Bezrizikovú úrokovú mieru sme použili zo stránky http://bonds.yahoo.com v hodnote *r = 0,0408*. Aktuálna spotová cena akcie bola dňa 23.8.2005 *S_0 = 43,44 USD*. Čas vypršania opcie je sobota po treťom piatku v mesiaci marec 2006, t.j *18. marca 2006*, z čoho pri aplikovaní štandardu *ACT/365* je *T = 0,3151*.

[1] Zdôvodnenie je uvedené napr. v Hull, J. 2000. *Options, Futures, and Other Derivate Securities*. Prentice-Hall International, Inc..

Po dosadení do (6):

$$C^E = S_0.\Phi(d_1) - K.e^{-rT}\Phi(d_2) = 43,44.\Phi\left(\frac{\ln\frac{43,44}{40} + 0,0408.0,3151 + \frac{1}{2}.0,18414^2.0,3151}{0,18414.\sqrt{0,3151}}\right) -$$

$$-40.e^{-0,0408.0,3151}\Phi\left(\frac{\ln\frac{43,44}{40} + 0,0408.0,3151 - \frac{1}{2}0,18414^2.0,3151}{0,18414\sqrt{0,3151}}\right) =$$

$$= 43,44.\Phi(0,974) - 39,489.\Phi(0,871) = 43,44 \text{ . } 0,835 - 39,489 \text{ . } 0,808 = 4,37 \text{ USD.}$$

Následne môžeme využiť predajno-kúpnu paritu. Zo vzťahu (3) vyplýva:

$$P_0^K = K.e^{-rT} + C_0^K - S_0 = 40.e^{-0,0408.0,3151} + 4,37 - 43,44 = 39,489 + 4,37 - 43,44 = 0,419 \text{ USD.}$$

4. Záver

Pomocou predajno-kúpnej parity sme zistili, že spravodlivá cena pre budúcoročnú (2006) marcovú európsku predajnú opciu na akciu SAP je 0,419 USD. Kvalitu nášho výpočtu môžeme verifikovať reálnymi dátami uvádzanými 23. novembra 2005 o 19.10 h na stránke http://www.cboe.com :

Calls	Last Sale	Net	Bid	Ask	Vol	Open Int	Puts	Last Sale	Net	Bid	Ask	Vol	Open Int
06 Mar 37.50 (SAP CT-E)	6.80	+0.50	6.80	7.00	10	100	06 Mar 37.50 (SAP OT-E)	0.65	pc	0.25	0.35	0	253
06 Mar 40.00 (SAP CH-E)	4.20	pc	4.60	4.80	0	109	06 Mar 40.00 (SAP OH-E)	1.25	pc	0.55	0.70	0	157
06 Mar 42.50 (SAP CS-E)	2.50	pc	2.85	2.95	0	319	06 Mar 42.50 (SAP OS-E)	1.70	pc	1.20	1.35	0	291

Z výsledkov je zrejmé, že náš odhad je relatívne presný.

Zoznam použitej literatúry

[1] Black, F. – Scholes, M.: The Pricing of Options and Corporate Liabilities. In: Journal of Political Economy, 81 (1973), s. 87 – 99.
[2] Hull, J.C. 2000. *Options, Futures and Others Derivates*. Fourth edition. London : Prentice-Hall International 2000, 698 s. ISBN: 0-13-015822-4
[3] Chajdiak, J. – Komorník, J. – Komorníková, M. 1999. *Štatistické metódy*. Bratislava : Statis, 1999, 282 s. ISBN: 80-85659-13-1
[4] Melicherčík, I. – Olšárová, L. – Úradníček, V. 2005. *Kapitoly z finančnej matematiky 1*. Zvolen : Bratia Sabovci, s.r.o. 2005, 138 s. ISBN: 80-89029-88-4

Ing. Vladimír Úradníček, PhD.
Katedra kvantitatívnych metód a informatiky
Oddelenie štatistiky a ekonomickej analytiky
Ekonomická fakulta Univerzity Mateja Bela
Tajovského 10
975 90 Banská Bystrica
vladimir.uradnicek@umb.sk

Aplikácia neurónových sietí pri modelovaní agrárnej zamestnanosti na Slovensku

Zita Varacová

ABSTRAKT

Do roku 1989 malo v politike zamestnanosti významnú úlohu poľnohospodárstvo, kde potom zase nastalo značné znižovanie počtu zamestnancov. Cieľom práce je zanalyzovať vývoj zamestnanosti v agrosektore na Slovensku, a to celkovo aj za jednotlivé profesijné skupiny od roku 1992 do roku 2002. Našou snahou bolo urobiť prognózy počtov zamestnancov v poľnohospodárstve do roku 2007 pomocou neurónových sietí.

KĽÚČOVÉ SLOVÁ

agrárna zamestnanosť, neurónové siete, profesijné skupiny v poľnohospodárstve

ABSTRACT

Agriculture had an important task in the policy of employment till the year 1989. Since then a considerable decrease in the number of employees has been a typical feature of agriculture. The goal of this work is to analyse the development of employment in agrarian sector in Slovakia from the year 1992 to year 2002. We tried to make prognoses of the number of employees in agriculture till 2007 by the means of neural networks.

KEY WORDS

agrarian employment, neural networks, professional groups in agriculture

ÚVOD

Značný pokles zamestnanosti zaznamenalo Slovensko od roku 1989 v poľnohospodárstve. U poľnohospodárskych družstiev a štátnych podnikov je úbytok pracovníkov trvalý. U nových podnikateľských subjektov v prvovýrobe, spracovateľskom priemysle, podpornom servise a samosprávnych organizáciách sa početné stavy pracovníkov zvyšujú. Nezamestnanosťou boli najviac postihnuté oblasti s vysokým podielom vidieckeho obyvatelstva a so zaostalou infraštruktúrou. V porovnaní s inými odvetviami je intenzita poklesu zamestnanosti v poľnohospodárstve najvyššia.

Znižovanie agrárnej zamestnanosti vedúce k rýchlemu rastu agrárnej nezamestnanosti, súviselo so zmenou vlastníckych vzťahov, postupnými zmenami v štruktúre výroby a zmenou organizácie výroby a práce. Je to dôsledok transformácie národného hospodárstva na trhovú ekonomiku. Pre poľnohospodársku a potravinársku výrobu je charakteristická tiež previazanosť s ďalšími odvetviami hospodárstva a to aj napriek tomu, že v priebehu

transformácie došlo zo známych príčin k jej útlmu (zníženie dopytu po potravinách a následne poľnohospodárskych výrobkov, po vstupoch z iných odvetví aj v dôsledku cenovej disparity, prehlbovanie pasívneho salda zahraničného obchodu). Z tohoto útlmu vyplynulo i znižovanie počtu pracovníkov v poľnohospodárstve.

Skutočnosť, do akej miery sa na Slovensku podarí zosúladiť dopyt a ponuku na trhu práce je odpoveďou na otázku, či sa v dohľadnej dobe dočkáme postupného významnejšieho znižovania nezamestnanosti v národnom hospodárstve.

CIEĽ

Cieľom príspevku je analyzovať vývoj zamestnanosti v agrárnom sektore na Slovensku a pokúsiť sa o jej predikciu pomocou neurónových sietí..

METODIKA

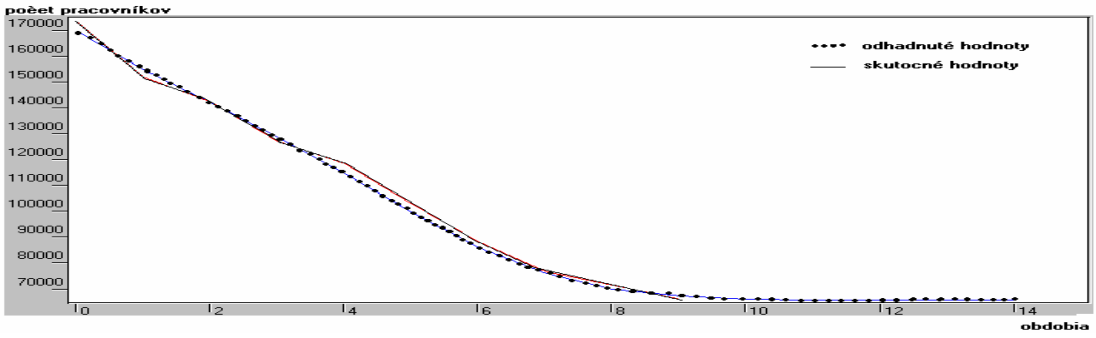
V práci sme sa zamerali na predikciu vývoja agrárnej zamestnanosti na 5 rokov pomocou metódy *neurónové siete*. Zaoberali sme sa nielen vývojom celkového počtu zamestnaných v poľnohospodárstve, ale aj vývojom počtu zamestnaných za jednotlivé profesijné skupiny, a to:

- trvale činní pracovníci (t.j. celkový počet zamestnancov v poľnohospodárstve),
- robotníci v rastlinnej a živočíšnej výrobe (t.j. traktoristi a mechanizační robotníci a iní robotníci v RV, ošetrovatelia zvierat),
- vedúci technickí a administratívni pracovníci,
- ostatní trvale činní zamestnanci.

Analýza bola vypracovaná na štatistických údajov zverejnených prostredníctvom ŠÚ SR, štatistických ročeniek a Štrukturálneho cenzu fariem SR.

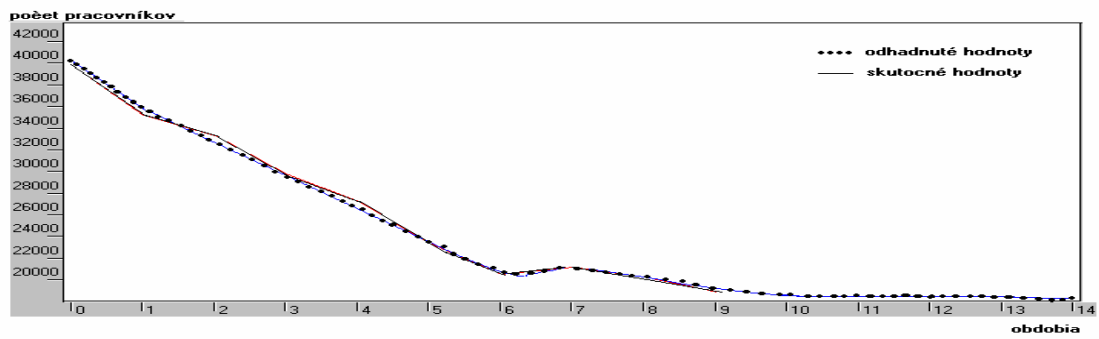
VÝSLEDKY A DISKUSIA

Graf 1 Vývoj počtu trvale činných pracovníkov v poľnohospodárstve



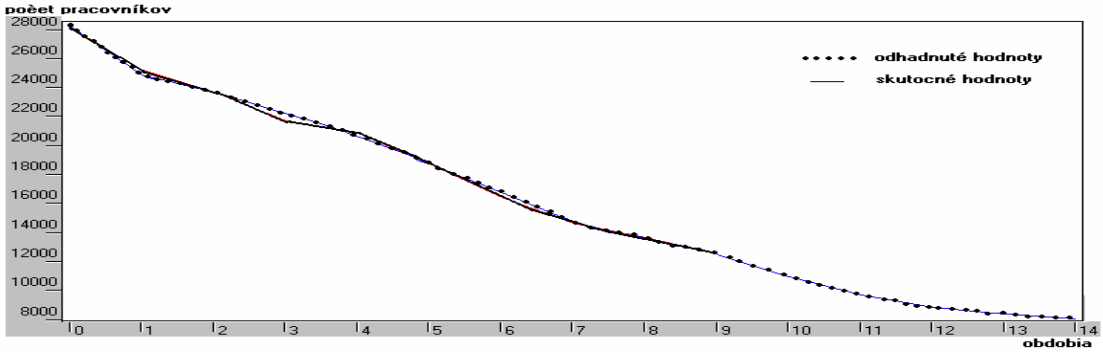
Prameň: [9-12] a vlastné výpočty

Graf 2 Vývoj počtu robotníkov v živočíšnej výrobe



Prameň [9-12] a vlastné výpočty

Graf 3 Vývoj počtu vedúcich technických a administratívnych zamestnancov



Prameň: [9-12] a vlastné výpočty

Vysvetlivky: THP – vedúci technickí a administratívni pracovníci

Popisy osi x:

0 – rok 1993	3 – rok 1996	6 – rok 1999	9 – rok 2002	12 – rok 2005
1 – rok 1994	4 – rok 1997	7 – rok 2000	10 – rok 2003	13 – rok 2006
2 – rok 1995	5 – rok 1998	8 – rok 2001	11 – rok 2004	14 – rok 2007

Výstupy programu STATISTICA Neural Networks 4.0 uvádzajú predošlé grafy.

Počet **trvale činných pracovníkov v poľnohospodárstve** má klesajúci charakter, i keď intenzita klesania sa od roku 2002 zmiernila a v 2007 by mal byť tento počet 64 000.

Počet **robotníkov v živočíšnej výrobe** klesol od roku 1992 do roku 2002 o 47,51 %, čo svedčí o úpadku živočíšnej výroby. Úroveň úžitkovosti a využívania reprodukčných vlastností zvierat je z chovateľského i ekonomického hľadiska neuspokojivá. Príčinami sú organizačné nedostatky, nedostatočná výživa a malá starostlivosť o jednotlivé fázy reprodukčného cyklu plemenníc. Nepriaznivý stav v úžitkovosti ošípaných často súvisí aj s nedostatkom financií na nákup kŕmnych zmesí.

Počet **vedúcich technických a administratívnych pracovníkov** klesol od roku 1993 do roku 2002 až o 44, 76 %. Pokles bol plynulý. Odchod skúsených pracovníkov podmienili nestabilnosť vonkajšieho ekonomického prostredia, sociálny tlak vonkajšieho prostredia, lepšie platobné podmienky v iných sektoroch, nízke spoločenské uznanie,

kultúrnosť práce, požiadavky, aby vo vedúcich funkciách rozhodovali vlastníci pôdy a podielníci.

ZÁVER

V minulosti aj v súčasnosti sa pre predpovede používajú klasické štatistické metódy. Mnohé z nich však majú za predpokladu korektného použitia obmedzené využitie. Cieľom práce bolo prezentovať nový netradičný prístup, neurónové siete, na predikciu vývoja počtu zamestnancov v poľnohospodárstve za profesijné skupiny za roky 2003-2007. Táto metóda sa pri modelovaní ukázala ako efektívna. V budúcnosti máme preto v záujme vyskúšať aplikovať na analýzu agrárnej zamestnanosti *fuzzy neurónové siete. Tento článok vznikol v rámci riešenia projektu KEGA 3/2060/04.*

LITERATÚRA

[1] BIELIK, P. a i. 2001. *Podnikovo hospodárska teória agrokomplexu*. 2. vyd. Nitra : Slovenská poľnohospodárska univerzita. 2001, 270 s. ISBN 80-7137-861-5.

[2] BOŽÍK, M. 1999. Postavenie agropotravinárskeho komplexu v národnom hospodárstve. In: *Agropotravinársky sektor v transformujúcej sa ekonomike*, Bratislava, 1999, VÚEPP

[3] PACÁKOVÁ, V. – RUBLÍKOVÁ, E. 1997. Analýza a prognóza vývoja ukazovateľov nezamestnanosti. In: *Slovenská štatistika a demografia*, b.v., 1997, č. 1, s. 18–31.

[4] *Pôdohospodárstvo a životné prostredie: Štruktúra zamestnancov v poľnohospodárstve v roku 1997.* Bratislava : ŠÚ SR. 1998, 81 s.

[5] *Pôdohospodárstvo a životné prostredie: Štruktúra zamestnancov v poľnohospodárstve v roku 2001. Bratislava :* ŠÚ SR. 2002, 83 s.

[6] *Poľnohospodárstvo v Slovenskej republike: Vybrané ukazovatele v rokoch 1970-2002.* Bratislava : ŠÚ SR. 2003, 86.

[7] STEHLÍKOVÁ, B. 2004. Nový metodologický prístup k prognózovaniu demografických časových radov. In: *Acta oeconomica et informatica*, roč. 7, 2004, č. 2, s. 10-11.

[8] ŠIMOVÁ, E. 1996. Analýza početných stavov pracovníkov v procese transformácie v poľnohospodárstve SR. In: *Acta operativo oeconomica*, Nitra : Vysoká škola poľnohospodárska, 1996, s. 87–90.

[9] *Štatistické čísla a grafy : Štruktúra pracovníkov v poľnohospodárstve v roku 1993 – 1997*. Bratislava : ŠÚ SR. 1998, 65 s.

[10] *Štatistická ročenka Slovenskej republiky*. Bratislava : SAV. 1991 - 2003.

[11] *Štrukturálny cenzus fariem 2001 : Fyzické osoby.* Bratislava : ŠÚ SR, 2002

[12] *Zelená správa: Súhrnná správa o poľnohospodárstve a potravinárstve v Slovenskej republike v roku 1997.* Bratislava : MP SR. 1998, 256 s.

[13] Základy neurónových sietí. 2003, [cit. 2003-06-12]. Dostupné na internete: <http://www.ii.fmph.uniba.sk/~farkas/Courses/ns.html>.

Autor:

Ing. Varacová Zita, Katedra štatistiky a operačného výskumu FEM SPU Nitra, zita.varacova@centrum.sk

Aplikácia zhlukovej analýzy fuzzy c – priemerov pri analýze agrárnej zamestnanosti na Slovensku

Zita Varacová

ABSTRAKT

Poľnohospodárstvo plnilo významnú úlohu v zamestnávaní obyvateľstva Slovenska až do obdobia transformácie. Cieľom práce je zanalyzovať vývoj zamestnanosti v agrosektore na Slovensku, a to celkovo aj za jednotlivé profesijné skupiny od roku 1992 do roku 2002 a nájsť významné etapy pre túto veličinu.

KĽÚČOVÉ SLOVÁ

agrárna zamestnanosť, zhluková analýza fuzzy c - priemerov, profesijné skupiny v poľnohospodárstve

ABSTRACT

Agriculture had an important task in the policy of employment in Slovakia till the transformation process. The goal of this work is to analyse the development of employment in agrarian sector in Slovakia from the year 1992 to year 2002 and to find significant periods for agricultural employment.

KEY WORDS

agrarian employment, fuzzy c –means cluster analysis, professional groups in agriculture

ÚVOD

Na začiatku transformácie nezamestnanosť nebola takmer vôbec známa, dnes patrí medzi akútne problémy hospodársko-sociálnej politiky. Príčinou rastu nezamestnanosti bol pokles dynamiky rastu a úpadok slovenského hospodárstva, štruktúrny regionálny a demografický vývoj. Značný pokles zamestnanosti sme zaznamenali v poľnohospodárstve.

Zhlukovou analýzou sme mali v záujme zistiť významné etapy vo vývoji zamestnanosti v poľnohospodárstve. Pod tvorbou zhlukov (etáp rokov) rozumieme proces vydelenia rokov, ktoré majú určité znaky a ich oddelenie od rokov, ktoré ich nemajú. Najprv sme zhlukovali roky pre premennú počet zamestnancov v poľnohospodárstve za jednotlivé profesné skupiny. Následne sme zhlukovali roky aj za premenné rastlinná produkcia, osevná plocha a počet poľnohospodárskych zvierat. Tým sme si chceli overiť súvis medzi vývojom počtu pracovníkov za profesie a vývojom orientácie poľnohospodárskej činnosti na rastlinnú a živočíšnu produkciu.

CIEĽ

Cieľom práce je nájsť etapy rokov, ktoré boli významné pre vývoj zamestnanosti v poľnohospodárstve na Slovensku.

METODIKA

V práci sme sa zaoberali nielen vývojom celkového počtu zamestnaných v poľnohospodárstve, ale aj vývojom počtu zamestnaných za jednotlivé profesijné skupiny, a to:

- trvale činní pracovníci (t.j. celkový počet zamestnancov v poľnohospodárstve),
- robotníci v rastlinnej a živočíšnej výrobe (t.j. traktoristi a mechanizační robotníci a iní robotníci v RV, ošetrovatelia zvierat),
- vedúci technickí a administratívni pracovníci,
- ostatní trvale činní zamestnanci.

Použitou metódou bola **zhluková analýza fuzzy c – priemerov** pre zistenie významných etáp a teda aj zlomových období vo vývoji zamestnanosti v poľnohospodárstve (v súvislosti s tým aj etáp vo vývoji rastlinnej produkcie a osevných plôch za významné skupiny plodín a etáp vo vývoji živočíšnej produkcie).

Analýza bola vypracovaná na základe štatistických údajov zverejnených prostredníctvom ŠÚ SR, štatistických ročeniek a Štrukturálneho cenzu fariem SR.

VÝSLEDKY A DISKUSIA

Pri každej premennej sme zhlukovali roky. Pri premennej počet zamestnancov nám išlo o zistenie dôležitých vývojových etáp počtu zamestnancov za jednotlivé profesie. Koeficienty celkovej separácie pre jednotlivé počty zhlukov uvádzame v tabuľke 1.

Tabuľka 1 Koeficienty celkovej separácie pre jednotlivé počty zhlukov

	2 zhluky	**3 zhluky**	**4 zhluky**	**5 zhlukov**	**6 zhlukov**	**7 zhlukov**	**8 zhlukov**	**9 zhlukov**	**10 zhlukov**
Koeficient separácie	9,8121	8,3784	10,1523	**10,7718 max**	10,2600	10,2381	9,6216	9,2081	9,0395

Prameň: vlastné výpočty

O optimálnom počte zhlukov sme sa rozhodli na základe hodnôt koeficientov celkovej separácie. Optimálny počet zhlukov je ten, pre ktorý je koeficient celkovej separácie maximálny. Výsledkom zhlukovej analýzy fuzzy c – priemerov je teda 5 etáp (zhlukov). Výsledné etapy rokov sú uvedené v tabuľke 2. Hodnoty funkcie príslušnosti pre týchto 5 etáp uvádza tabuľka 3.

Tabuľka 2 Etapy rokov pre profesijné skupiny zamestnancov v poľnohospodárstve

	1. etapa	2. etapa	3. etapa	4. etapa	5. etapa
Rok	1992	1993 1994 1995 1996	1997 1998 1999	2000	2001 2002

Prameň: vlastné výpočty

Tabuľka 3 Hodnoty funkcie príslušnosti

Rok	Zhluky				
1992	**1.0000**	0.0000	0.0000	0.0000	0.0000
1993	0.0016	0.0082	0.0045	**0.9777**	0.0079
1994	0.0001	0.0007	0.0004	**0.9982**	0.0006
1995	0.0002	0.0013	0.0006	**0.9967**	0.0012
1996	0.0014	0.0114	0.0050	**0.9724**	0.0099
1997	0.0012	0.0092	0.0044	0.0075	**0.9777**
1998	0.0002	0.0014	0.0007	0.0011	**0.9967**
1999	0.0012	0.0113	0.0058	0.0079	**0.9739**
2000	0.0000	**1.0000**	0.0000	0.0000	0.0000
2001	0.0002	0.0016	**0.9957**	0.0011	0.0013
2002	0.0003	0.0018	**0.9953**	0.0012	0.0014

Prameň: vlastné výpočty

Každý z rokov s hodnotou funkcie príslušnosti vyššou ako 0,80 nazveme vyhraneným. Ostatné sú nevyhranené. Rok s hodnotou funkcie príslušnosti vyššou ako 0,80 patriaceho do daného zhluku nazveme predstaviteľom tohoto zhluku. Nevyhranené roky nazveme inklinujúcimi k tomu zhluku, ktorého funkcia príslušnosti je maximálna. Predstaviteľom prvej etapy (zhluku) je rok 1992 (hodnota funkcie príslušnosti 1,0000), predstaviteľmi druhej etapy sú roky 1993 (0.9777), 1994 (0,9982), 1995 (0,9967) a 1996 (0,9724), predstaviteľmi tretej etapy sú roky 1997 (0,9777), 1998 (0,9967) a 1999 (0,9739), predstaviteľom štvrtej etapy je rok 2000 (1,0000) a predstaviteľmi piateho zhluku sú roky 2001 (0,9957) a 2002 (0,9953). Nevyhranené roky v našom prípade neexistujú.

ZÁVER

Z výsledkov zhlukovania možno konštatovať, že sa nám vytvorili etapy rokov typické pre vývoj počtu zamestnancov v poľnohospodárstve, ale aj rastlinnej produkcie, osevnej plochy a tiež aj pre počty poľnohospodárskych zvierat. Po podrobnom posúdení etáp za všetky premenné sme zistili, že významné obdobie pre vývoj počtu zamestnancov v poľnohospodárstve za jednotlivé profesie sú roky 1992, 1993 – 1996, 1997 – 1999, 2000 a 2001-2002. Pre osevnú plochu sa vytvorili etapy rokov 1992, 1993-1994, 1995-1998, 1999 a 2000-2002. Etapa rokov 1997 – 1998 sa ukázala ako významná aj pre osevnú plochu, rastlinnú produkciu aj počty zvierat. Potvrdila sa nám hypotéza, že pre vývoj počtu

pracovníkov v poľnohospodárstve za profesné skupiny a pre oblasť rastlinnej a živočíšnej produkcie sa nám vytvorila etapa rokov, ktorá je pre ne spoločná. Išlo najmä o roky 1997-1999.

LITERATÚRA

[1] BIELIK, P. a i. 2001. *Podnikovo hospodárska teória agrokomplexu*. 2. vyd. Nitra : Slovenská poľnohospodárska univerzita. 2001, 270 s. ISBN 80-7137-861-5.

[2] BOŽÍK, M. 1999. Postavenie agropotravinárskeho komplexu v národnom hospodárstve. In: *Agropotravinársky sektor v transformujúcej sa ekonomike*, Bratislava, 1999, VÚEPP

[3] PACÁKOVÁ, V. – RUBLÍKOVÁ, E. 1997. Analýza a prognóza vývoja ukazovateľov nezamestnanosti. In: *Slovenská štatistika a demografia*, b.v., 1997, č. 1, s. 18–31.

[4] *Pôdohospodárstvo a životné prostredie: Štruktúra zamestnancov v poľnohospodárstve v roku 1997*. Bratislava : ŠÚ SR. 1998, 81 s.

[5] *Pôdohospodárstvo a životné prostredie: Štruktúra zamestnancov v poľnohospodárstve v roku 2001. Bratislava :* ŠÚ SR. 2002, 83 s.

[6] *Poľnohospodárstvo v Slovenskej republike: Vybrané ukazovatele v rokoch 1970-2002.* Bratislava : ŠÚ SR. 2003, 86.

[7] STEHLÍKOVÁ, B. 2004. Nový metodologický prístup k prognózovaniu demografických časových radov. In: *Acta oeconomica et informatica*, roč. 7, 2004, č. 2, s. 10-11.

[8] ŠIMOVÁ, E. 1996. Analýza početných stavov pracovníkov v procese transformácie v poľnohospodárstve SR. In: *Acta operativo oeconomica*, Nitra : Vysoká škola poľnohospodárska, 1996, s. 87–90.

[9] *Štatistické čísla a grafy : Štruktúra pracovníkov v poľnohospodárstve v roku 1993 – 1997*. Bratislava : ŠÚ SR. 1998, 65 s.

[10] *Štatistická ročenka Slovenskej republiky.* Bratislava : SAV. 1991 - 2003.

[11] *Štrukturálny cenzus fariem 2001 : Fyzické osoby.* Bratislava : ŠÚ SR, 2002

[12] *Zelená správa: Súhrnná správa o poľnohospodárstve a potravinárstve v Slovenskej republike v roku 1997*. Bratislava : MP SR. 1998, 256 s.

Autor:

Ing. Varacová Zita, Katedra štatistiky a operačného výskumu FEM SPU Nitra, zita.varacova@centrum.sk

Použitie kanonickej korelačnej analýzy pri riešení závislosti ukazovateľov veľkosti potravinárskych priemyselných podnikov SR

Mária Vojtková

Abstract

In this article we illustrate the use canonical correlation analysis to analyze the relationship between two sets of variables. The sources of information are the output and the input variables of the size of enterprises on industry of the Slovak Republic in the year 2003.

Úvod

Kanonická korelačná analýza je viacrozmerná štatistická metóda, ktorá sa nachádza na rozhraní medzi faktorovou analýzou a viacnásobnou regresnou a korelačnou analýzou. Jej cieľom je, čo najjednoduchšie a zároveň najvýstižnejšie opísať závislosť medzi dvomi skupinami pôvodných premenných pomocou ich lineárnych kombinácií.

Jej použitie si vyžaduje predovšetkým tá skutočnosť, že pri tejto analýze skúmame závislosť medzi väčšou skupinou závislých premenných Y_1, Y_2, ..., Y_m a podobne ako pri viacnásobnej regresnej analýze skupinou nezávislých premenných X_1, X_2, ..., X_k , avšak na rozdiel od spomínaných metód z oboch skupín sú postupne určované dvojice lineárnych kombinácií premenných za prvú aj druhú skupinu tak, aby medzi týmito dvojicami bola maximálna korelácia.

Údajová základňa

Riešenie problému závislosti medzi skupinami premenných sme aplikovali na ukazovatele charakterizujúce veľkosť priemyselných podnikov 15. odielu Odvetvovej klasifikácie ekonomických činností (OKEČ) „Výroba potravín a nápojov“ na Slovensku v roku 2003. Ide o anonymizované podnikové údaje, poskytnuté pre vedecké účely zo Štatistického úradu SR spracovaním predbežných ročných údajov.

Zaujímal nás vzťah medzi vybranými ukazovateľmi charakterizujúcimi výstup výrobného procesu v potravinárskom podniku (Y_1, Y_2, Y_3):

PH - objem pridanej hodnoty v tis. Sk,
HV - výsledok hospodárenia pred zdanením v tis. Sk,
V - objem výnosov v tis. Sk,

a vybranými ukazovateľmi charakterizujúcimi vstup do výrobného procesu (X_1, X_2, X_3):

CK - stav celkového kapitálu (pasíva spolu) v tis. Sk,
ZAS - stav zásob v tis. Sk,
ZAM - priemerný evidenčný počet zamestnancov v prepočítaných osobách.

Výsledky kanonickej korelačnej analýzy

Na úvod sme overili štatistickú významnosť kanonickej korelačnej analýzy pomocou viacerých aproximatívnych F – testov založených na viacrozmerných štatistikách. V našom prípade všetky viacrozmerné štatistiky a na to naväzujúce aproximatívne F-testy potvrdili, že medzi oboma skupinami premenných existuje štatisticky významná závislosť. Na základe p-hodnoty na ľubovolnej hladine významnosti vyššej ako 0.0001 zamietame hypotézu H_0 o nezávislosti skupinového koeficienta kanonickej korelácie.

Multivariate Statistics and F Approximations					
S=3 M=-0.5 N=144.5					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.04918815	192.64	9	708.37	<.0001
Pillai's Trace	1.28474219	73.15	9	879	<.0001
Hotelling-Lawley Trace	12.68235718	409.05	9	454.11	<.0001
Roy's Greatest Root	12.14768441	1186.42	3	293	<.0001

Obr. 1 Viacrozmerné štatistiky použité pri teste významnosti skupinového koeficienta kanonickej korelácie

Ďalej sme sa zamerali na overenie štatistickej významnosti jednotlivých dvojíc kanonických premenných modifikáciou F-testu a tiež hodnoty jednotlivých koeficientov kanonickej korelácie charakterizujúcich závislosťmedzi dvojicami.

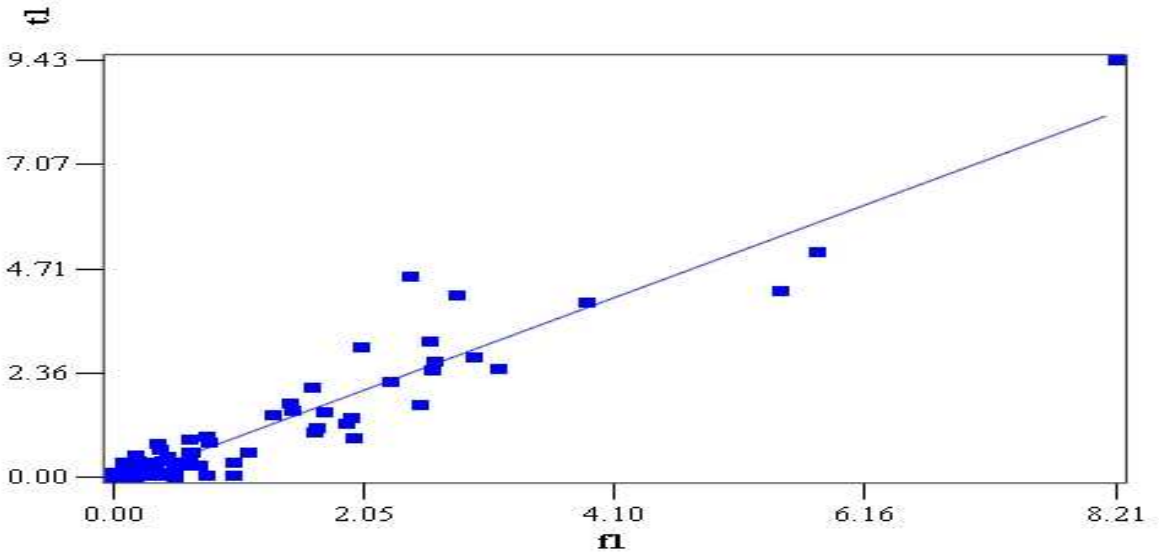
	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation	Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)			
					Eigenvalue	Difference	Proportion	Cumulative
1	0.961218	0.960795	0.004421	0.923941	12.1477	11.6357	0.9578	0.9578
2	0.581909	0.578711	0.038442	0.338619	0.5120	0.4893	0.0404	0.9982
3	0.148938	.	0.056834	0.022183	0.0227		0.0018	1.0000

	Test of H0: The canonical correlations in the current row and all that follow are zero				
	Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
1	0.04918815	192.64	9	708.37	<.0001
2	0.64671022	35.55	4	584	<.0001
3	0.97781742	6.65	1	293	0.0104

Obr. 2 Kanonické korelačné koeficienty pre jednotlivé dvojice kanonických premenných a testy ich významnosti

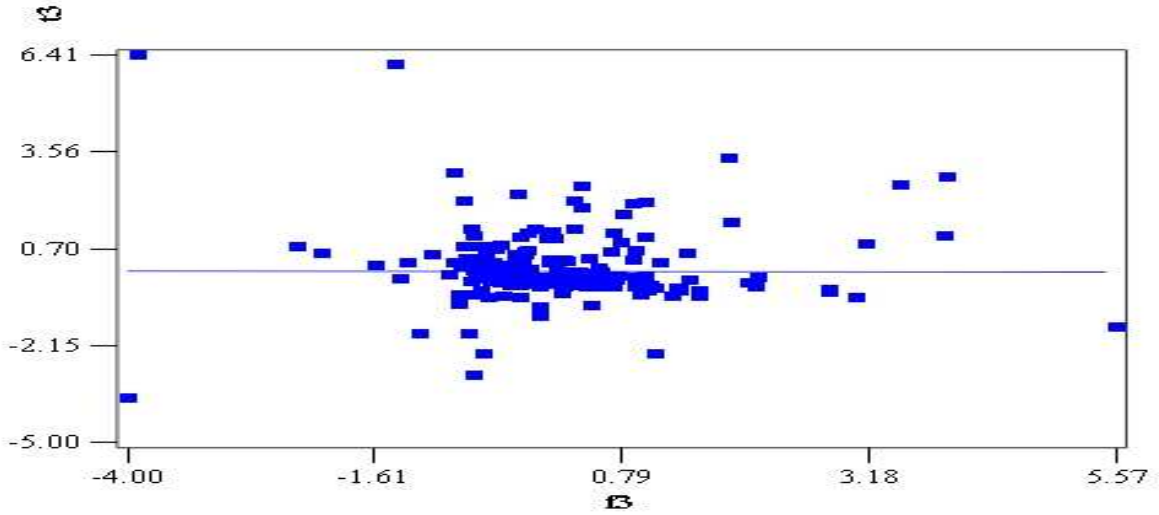
Hodnoty koeficientov kanonickej korelácie sú zoradené podľa veľkosti, čiže prvá dvojica kanonických premenných dosahuje najvyššiu závislosť 0,961218, druhá dvojica strednú závislosť 0,581909 a pri poslednej dvojici je hodnota koeficienta kanonickej korelácie nízka 0,148938. Už na základe týchto hodnôt, ale tiež podľa p-hodnoty uvedenej v riadkoch pre jednotlivé koeficienty kanonickej korelácie je možné zamietnuť nulovú hypotézu o nevýznamnosti prvých dvoch koeficientov . Posledný koeficient kanonickej korelácie je na hladine 0,01 štatisticky nevýznamný a preto s ním v ďalšej interpretácii nebudeme uvažovať.

Na ukážku sme vytvorili graf závislosti pre prvú dvojicu kanonických premenných, ktorú považujeme na základe predchádzajúcich analýz za štatisticky významnú a maximálne korelovanú.



Obr. 3 Grafické znázornenie závislosti medzi prvou dvojicou kanonických premenných

Na porovnanie môžeme urobiť graf i pre tretiu dvojicu, ktorá bola zhodnotená ako štatisticky nevýznamná. Tento graf iba potvrdzuje predchádzajúce zistenia.



Obr. 4 Grafické znázornenie závislosti medzi treťou dvojicou kanonických premenných

Standardized Canonical Coefficients for the Set 1 Variables				
		t1	**t2**	**t3**
PH	PH	0.6188	-0.5809	-1.6344
HV	Zpred	-0.1571	1.0051	-0.2460
V	V	0.4598	0.4389	1.7085

Standardized Canonical Coefficients for the Set 2 Variables				
		f1	**f2**	**f3**
ZAS	ZAS	-0.1028	1.6117	0.3787
ZAM	ZAM	0.3172	0.0548	1.4693
CK	CK	0.8201	-1.1900	-1.5796

Obr. 5 Štandardizované kanonické koeficienty pre obe skupiny vstupných premenných

Z predchádzajúceho obrázku vidíme, že prvá dvojica kanonických premenných má tvar:

$t_1 = 0{,}6188*PH - 0{,}1571*HV + 0{,}4598*V$
$f_1 = -0{,}1028*ZAS + 0{,}3172*ZAM + 0{,}8201*CK$

a závislosť medzi touto dvojicou kanonických premenných je vysoká $r_c = 0{,}96$. Podľa hodnôt váh jednotlivých kanonických premenných, môžeme zhodnotiť, že na prvú kanonickú premennú t_1, ktorá vznikla lineárnou kombináciou ukazovateľov charakterizujúcich výstup výrobného procesu má najväčší vplyv ukazovateľ PH a na prvú kanonickú premennú f_1, ktorá vznikla lineárnou kombináciou ukazovateľov charakterizujúcich vstup výrobného procesu má vplyv predovšetkým ukazovateľ CK.

Podobne môžeme interpretovať aj druhú štatisticky významnú dvojicu:

$t_2 = -0{,}5809*PH - 1{,}0051*HV + 0{,}4389*V$
$f_2 = 1{,}6117*ZAS + 0{,}0548*ZAM - 1{,}1900*CK$

kde sa prejavuje predovšetkým rozhodujúci vplyv premennej HV na prvú z druhej dvojice kanonických premenných a ukazovateľa ZAS na druhú z uvedenej dvojice.

Uvedenú interpretáciu si upresníme a porovnáme s veľkosťou koeficientov korelácie medzi pôvodnými a kanonickými premennými.

Correlations Between the Set 1 Variables and Their Canonical Variables				
		t1	**T2**	**t3**
PH	PH	0.9570	0.0814	-0.2784
HV	Zpred	0.1443	0.9484	-0.2824
V	V	0.9363	0.2145	0.2783

Correlations Between the Set 2 Variables and Their Canonical Variables				
		f1	**f2**	**f3**
ZAS	ZAS	0.6870	0.7053	-0.1746
ZAM	ZAM	0.8662	-0.0613	0.4959
CK	CK	0.9704	0.1121	-0.2137

Obr. 6 Koeficienty korelácie medzi jednotlivými kanonickými a vstupnými premennými

Úplná interpretácia kanonických premenných je vo viacerých prípadoch dosť obtiažna. Z prvého páru kanonických premenných usudzujeme, že lineárna kombinácia objemu pridanej hodnoty a výnosov silno priamo korelujú so stavom celkového kapitálu. Pri druhom páre výsledok hospodárenia priamo koreluje so stavom zásob.

Ďalšia analýza je zameraná na analýzu koeficientov determinácie a redundancie, čiže budeme skúmať podiel vysvetlenej variability jednotlivými kanonickými premennými.

Prvá kanonická premenná t_1 vysvetľuje 60,44 % variability premenných v 1. skupine ukazovateľov charakterizujúcich výstup výrobného procesu a druhá kanonická premenná t_2 31,73 % variability. Obe štatisticky významné kanonické premenné teda vysvetľujú v 1. skupine spolu 92,18 % variability, čo je dostatočne veľká časť celkovej variability.

Podobne môžeme interpretovať aj podiel variability v 2. skupine ukazovateľov charakterizujúcich vstup výrobného procesu pomocou ich vlastných lineárnych kombinácií. Prvou kanonickou premennou f_1 je vysvetlených 72,14 % variability a druhou kanonickou

premennou f_2 17,13 % variability, čo dokopy tvorí 89,26 % celkovej vysvetlenej variability premenných 2. skupiny.

Pri prvom páre kanonických premenných je podiel variability lineárnych kombinácií premenných 1. skupiny vysvetlený lineárnou kombináciou premenných 2. skupiny dosť veľky, t.j. 92,39 %. Pri druhom páre kanonických premenných je tento podiel vysvetlenej variability už iba 33,86 % a tretí pár vysvetľuje štatisticky nevýznamnú časť variability.

Standardized Variance of the Set 1 Variables Explained by					
Canonical Variable Number	Their Own Canonical Variables		Canonical R-Square	The Opposite Canonical Variables	
	Proportion	Cumulative Proportion		Proportion	Cumulative Proportion
1	0.6044	0.6044	0.9239	0.5584	0.5584
2	0.3173	0.9218	0.3386	0.1075	0.6659
3	0.0782	1.0000	0.0222	0.0017	0.6676

Standardized Variance of the Set 2 Variables Explained by					
Canonical Variable Number	Their Own Canonical Variables		Canonical R-Square	The Opposite Canonical Variables	
	Proportion	Cumulative Proportion		Proportion	Cumulative Proportion
1	0.7214	0.7214	0.9239	0.6665	0.6665
2	0.1713	0.8926	0.3386	0.0580	0.7245
3	0.1074	1.0000	0.0222	0.0024	0.7269

Obr. 7 Analýza determinácie a redundancie pre obe skupiny vstupných premenných

Analýza redundancie je obsiahnutá v posledných dvoch stĺpcoch oboch výstupov. Celková redundancia pre premenné 1. skupiny charakterizujúce výstup výrobného procesu je 0,6676. To znamená, že premenné druhej skupiny vysvetľujú na 66,76 % variabilitu premenných v prvej skupine, čiže informácie o premenných prvej skupiny sú na 33,24 % prebytočné alebo redundantné. Avšak najviac sa na vysvetlení tejto variability podieľa prvá kanonická premenná f_1, ktorá vysvetľuje až 55,84 % variability a druhá kanonická premenná f_2 už iba 10,75 %.

Celková redundancia pre premenné 2. skupiny charakterizujúce vstup výrobného procesu je o niečo vyššia 0,7269, čo signalizuje 27,31% prebytočnosť informáciií o tejto skupine ak máme k dispozícií premenné prvej skupiny. Prvá kanonická premenná $t_1$vysvetľuje 66,65 % variability a druhá kanonická premenná t_2 vysvetľuje 5,8 % variability premenných v druhej skupine.

Záver

Uvedená analýza poukazuje na možnosť riešenia problému závislosti v prípade existencie väčšej skupiny nezávislých premenných pomocou kanonickej korelácie. Pričom analýza konkrétneho súboru ukazovateľov potvrdila v našom prípade opodstatnennosť existencie dvoch párov kanonických premenných, čo viedlo k zníženiu dimenzie pôvodného

súboru ukazovateľov. Následne je možná analýza už zredukovaného súboru prostredníctvom ďalších i jednoduchších metód.

Literatúra

[1] DILLON, R. W. – GOLDSTEIN, M.: *Multivariate Analysis: Methods and applications*. New York: John Wiley & Sons, 1984.
[2] HAIR, J. F. – ANDERSON, R. E. – TATHAM, R. L. – BLACK, W. C.: *Multivariate data analysis.* New York: Macmillan Publishing Company, 1992.
[3] HEBÁK, P. - HUSTOPECKÝ, J. - JAROŠOVÁ, E. – PECÁKOVÁ, I.: *Vícerozmìrné statistické metódy (1).* Informatorium, Praha 2004.
[4] CHAJDIAK, J.: *Analýza stavu ekonomiky podnikov a odvetví*. Bratislava: Statis, 1999.
[5] LUHA J.: Viacrozmerné štatistické metódy analýzy kvalitatívnych znakov. Štatistické metódy v praxi. *EKOMSTAT 2005,* SŠDS Trenčianske Teplice 22. – 27. 5. 2005.
[6] SHARMA, S.: *Applied multivariate techniques*. New York: John Wiley & Sons, 1996.
[7] STANKOVIČOVÁ I.: *Viacrozmerná analýza rentability poisťovní SR pomocou Enterprise Guide.* Zborník príspevkov. Výpočtová štatistika 2001. 10. medzinárodný seminár, s. 110-114, Bratislava: SŠDS, 2001. ISBN80-88946-14-X

Kontakt:

Ing. Mária Vojtková, PhD.
Katedra štatistiky, FHI, EU Bratislava
E-mail: vojtkova@euba.sk

Starnutie

Ladislava Wsólová, Alexandra Horská

Abstract
Why we age. We have tried to find the answer in our project THE ROLE OF GENETIC PREDISPOSITION, IMMUNE SYSTEM AND NUTRITION IN AGING. Aim of our proposal is to understand basic processes of physiological aging, provide a sound scientific basis for efforts to promote healthy aging and delay the onset of disability. We collected 292 healthy volunteers: 151 young persons, age from 18 to 25 years and 141 older persons, age from 63 to 70 years. The young group consisted of 73 men and 78 women, the old group consisted of 39 men and 102 women. We have asked which variables associate with DNA or chromosome damage. Results: Single strand breaks in lymphocyte DNA (sbs) were higher in old group compared with young one. The highest level of sbs was observed in the group of GSTT1 positive older subjects. The frequency of micronuclei (nmn) was higher in old group compared with young one, in women compared with men, and in persons with higher BMI (body mass index).

Od dávnych čias sa ľudia snažili nájsť prameň večnej mladosti. Ako žiť dlho a vyzerať mlado a zostať zdravý? Aj dnes nás fascinujú príbehy ľudí, ktorí údajne žijú 100 a viac rokov. Sú na svete ľudia, ktorí sa dožili 150 a viac rokov?

Starnutie v súčasnosti patrí medzi najatraktívnejšie problémy. Ak by sme skutočne dokázali oddialiť starnutie, ľudský život by sa dramaticky zmenil. Predstavme si, že by spolu v rodine žilo 6-8 generácií a že by sme sa mohli dožiť dôsledkov krátkozrakých politických pochabostí.

Ak chceme skúmať príčiny starnutia, musíme byť schopní starnutie definovať a merať ho. Je ťažké merať starnutie u jednotlivca, ale je možné merať rýchlosť starnutia populácie. Všeobecne možno povedať, že pravdepodobnosť úmrtia má určitú hodnotu pri narodení, postupne klesá na nejakú nízku hodnotu a potom začína stúpať. A stúpa geometrickým radom po zvyšok života. Rýchlosť, s akou rastie pravdepodobnosť úmrtia s pribúdajúcimi rokmi, je dobrou mierou rýchlosti starnutia.

Náš život sa stal v priebehu stáročí bezpečnejším vďaka mnohým zdravotníckym opatreniam, ale napriek tomu nie sme schopní zaistiť, aby naše telesné schránky nechátrali.

Je starnutie genetické? Genetici zdôrazňujú, že pôsobenie génov a prostredia nemožno hodnotiť oddelene. Dokonca by sa dalo povedať, že na starnutí sa podieľajú všetky gény jedinca - genofond získaný od rodičov a aktuálny stav genómu, ktorý odráža chyby pri replikácii DNA, vplyvy životného štýlu a environmentálnych faktorov.

Existujú choroby ako napr. progéria alebo Downov syndróm, ktoré imitujú niektoré znaky starnutia. Ale **starnutie je všeobecné chátranie, počas ktorého sa veľa problémov vyskytuje stále častejšie**.

Ženy umierajú nižšou rýchlosťou a preto žijú dlhšie, ale nestarnú pomalšie. Jednoducho zomierajú nižšou rýchlosťou od počatia až po kremáciu. Sú biologicky silnejšie. Ale prečo?

Výskumníci majú nástroje na výskum špecifických ochorení zviazaných so starnutím. Avšak špecifické choroby nerovnajú sa starnutiu.

Prečo starneme?

Odpoveď hľadáme aj v rámci projektu ÚLOHA GENETICKEJ PREDISPOZÍCIE, IMUNITNÉHO SYSTÉMU A VÝŽIVY V PROCESE STARNUTIA.

Naším cieľom je výskum procesov starnutia za použitia multidisciplinárneho a molekulárno-epidemiologického prístupu. Chceme skúmať vplyv výživy, životného štýlu, imunitnej odpovede organizmu a individuálnej predispozície jedinca (genetických polymorfizmov) na proces starnutia a ako sa to odzrkadlí na úrovni bunkových markerov. Chceme získať informácie o bunkových a molekulárnych základoch starnutia pomocou sledovania vhodných biomarkerov dvoch vekovo rozdielnych vyvážených skupín obyvateľstva. Genetické faktory, ktoré môžu modulovať starnutie, budeme skúmať pomocou identifikácie polymorfizmov v génoch kódujúcich systémy priamo či nepriamo zabezpečujúce stabilitu a integritu DNA. Sú to gény pre vybrané enzýmy biotransformácie exogénnych látok (xenobiotík) a proteíny reparácie DNA. Chceli by sme nájsť vzájomné vzťahy medzi molekulárnymi markermi, psychologickými a fyzickými aspektami starnutia, ako aj rozdiely v starnutí medzi pohlaviami.

Vieme, že genetické faktory, ktoré podmieňujú vnímavosť jedinca k určitým ochoreniam, vykazujú u rôznych populácií rozdiely vo frekvencii výskytu. K faktorom životného štýlu a prostredia sa zaraďujú riziká, ktoré zvyšujú možnosť výskytu určitých ochorení, ako napr. konzumácia alkoholu, choroby z povolania, preľudnenie, rozdiely v príjmoch a výživa. Jedna z teórií považuje za hlavnú príčinu starnutia akumuláciu poškodenia v biomolekulách, ktorá môže byť aj následkom nedostatočného príjmu antioxidantov vo výžive. Hromadenie predovšetkým oxidačného poškodenia v lipidoch a proteínoch ohrozuje najmä bunky v tkanivách s dlhou životnosťou. Oxidačné poškodenie DNA môže mať ešte závažnejšie následky, lebo keď sa mutácia zafixuje, je nezvratná a môže spôsobiť uhynutie alebo v horšom prípade modifikáciu vlastností bunky (napr. vznik nádorov). Poškodenie DNA je výsledkom účinku poškodzujúcich agens na jednej strane a rýchlosti reparačných procesov na strane druhej. Rozsah poškodenia limitujú obranné antioxidačné mechanizmy, napr. enzýmy ako kataláza a superoxiddismutáza, glutatiónperoxidáza, tripeptid glutatión, spolu s diétnymi antioxidantmi, ako sú napr. vitamín C, vitamín E, karotenoidy a mnohé ďalšie známe i neznáme molekuly.

V súčasnosti končíme zber a vkladanie údajov a začíname s ich analýzou. Máme súbor 292 zdravých dospelých, mladšiu skupinu tvorí 151 osôb vo veku od 18 do 25 rokov, staršiu 141 osôb vo veku od 63 do 70 rokov. Mladšia skupina pozostáva zo 73 mužov a 78 žien, staršia z 39 mužov a 102 žien.

Položili sme si otázku: s ktorými premennými je asociované poškodenie DNA a ako sa tieto vzťahy prejavujú na dvoch úrovniach odlišných v citlivosti detekcie poškodenia, závažnosti a zvratnosti zmien. Parameter sbs (jednovláknové zlomy DNA) meraný metódou Comet assay veľmi citlivo odráža aktuálny stav poškodenej DNA v bunke, môže sa v čase pomerne dynamicky meniť v závislosti od exo- i endogénnych vplyvov a reparačného systému bunky. Parameter nmn reprezentuje počet mikrojadier v lymfocytoch a vyjadruje konečný, už nemenný stav poškodenia DNA. Mikrojadro vzniká oddelením časti DNA z chromozómu a jej

ohraničením membránou. Mikrojadrá sú konečné, neopravujú sa a ich nadmerná akumulácia vedie k likvidácii bunky. Obidva druhy poškodenia sme sledovali na vitálnych lymfocytoch z venóznej krvi Na postavenie hypotéz sme použili modul AnswerTree. Ďalej sme pracovali v SPSS 13. 0. Každý typ poškodenia DNA charakterizuje jeden matematický model.

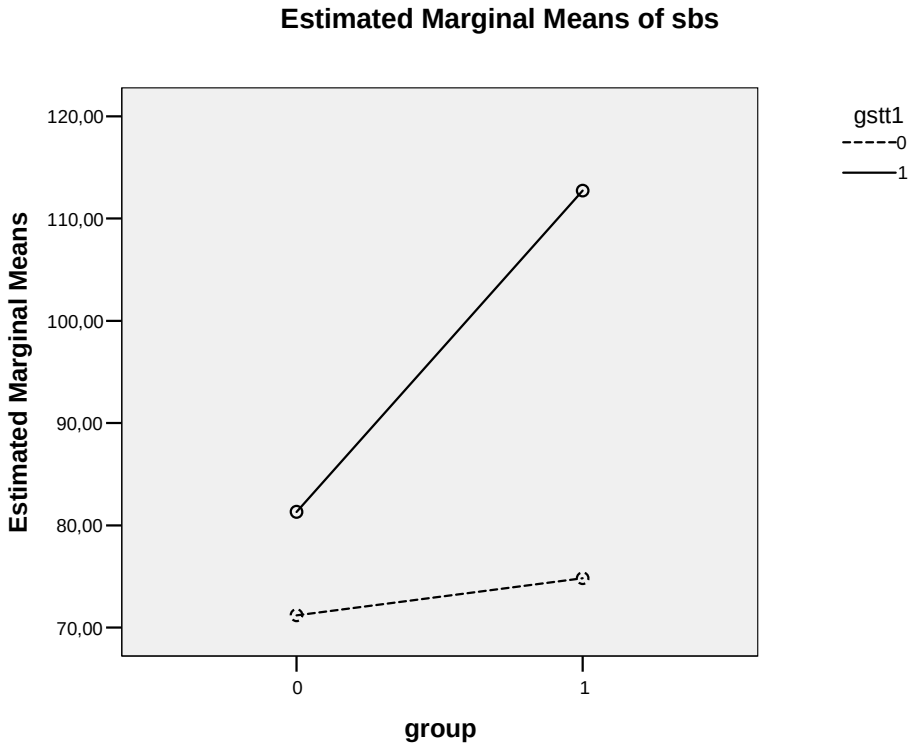
Tests of Between-Subjects Effects

Dependent Variable: sbs

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	59209,512[a]	3	19736,504	7,462	,000
Intercept	1872034,391	1	1872034,391	707,794	,000
group	19847,816	1	19847,816	7,504	,007
group * gstt1	47158,031	2	23579,016	8,915	,000
Error	698249,869	264	2644,886		
Total	2622481,990	268			
Corrected Total	757459,381	267			

a. R Squared = ,078 (Adjusted R Squared = ,068)

Závislá premenná sbs označuje aktuálne poškodenie DNA. Nezávislá premenná group znamená príslušnosť k vekovej skupine (0=mladší, 1= starší), gstt1 je označenie genotypu GSTT1 kódovaného číselne (0=osoba nemá tento gén, 1=osoba má tento gén). Model ukazuje, že starší ľudia majú priemerne vyššie aktuálne poškodenie DNA ako mladší. Signifikantné rozdiely v sbs medzi týmito dvomi vekovými skupinami prináša až väzba na GSTT1 (gén pre glutatiónтransferázu triedy theta, významný biotransformačný enzým). Prítomnosť tohto génu znevýhodňuje jeho nositeľov. Starší ľudia s chýbajúcim GSTT1 génom majú signifikantne nižšie aktuálne poškodenie DNA v porovnaní s jedincami s funkčným génom. Systém biotransformačných enzýmov metabolicky modifikuje exogénne látky v tele tak, aby mohli byť zneškodnené a následne vylúčené. Na vizualizáciu jeho vplyvu na organizmus je teda potrebný pomerne dlhý časový horizont, počas ktorého je jedinec vystavený pôsobeniu rozmanitých, vo všeobecnosti negatívnych vonkajších faktorov. Preto sa v skupine mladších tento vzťah ešte neprejavil. Prirodzene by sme očakávali, že funkčný variant GSTT1 bude pre ľudí benefitom. GSTT1 síce má výrazne ochrannú (detoxifikačnú a antioxidačnú) funkciu, ale bolo potvrdené, že niektoré exogénne molekuly GSTT1 práve aktivuje a zvyšuje ich toxicitu. Nami navrhnutý model naznačuje, že GSTT1 môže metabolicky aktivovať agens, ktoré selektívne poškodzujú DNA.



Tests of Between-Subjects Effects

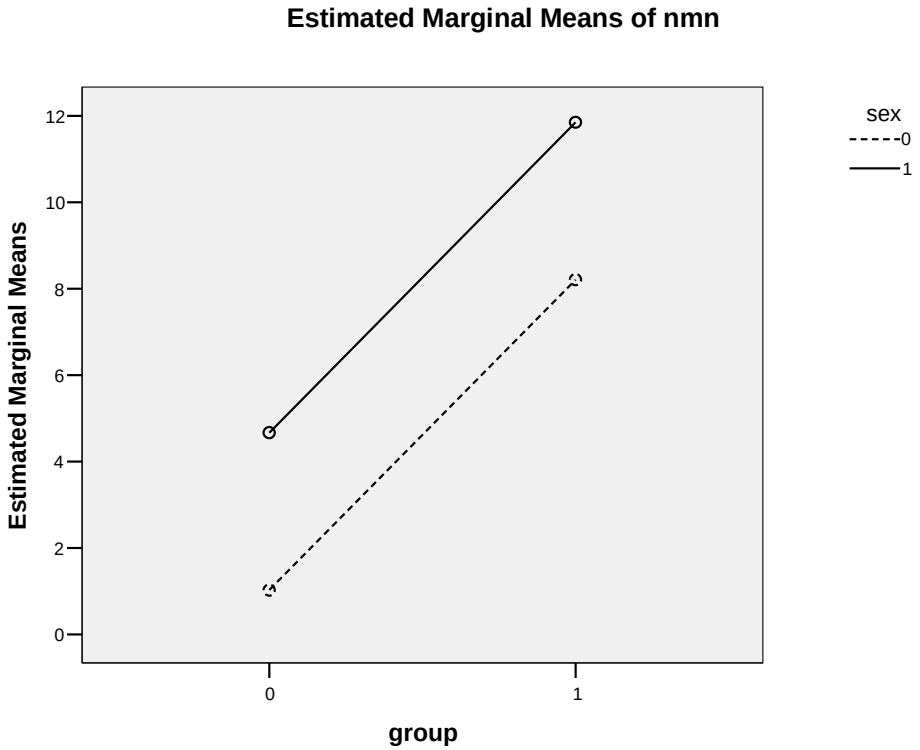
Dependent Variable: nmn

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	18022,390[a]	4	4505,597	135,682	,000
group	1478,211	1	1478,211	44,515	,000
BMI	277,301	1	277,301	8,351	,004
sex	640,028	1	640,028	19,274	,000
Error	7471,610	225	33,207		
Total	25494,000	229			

a. R Squared = ,707 (Adjusted R Squared = ,702)

Závislá premenná nmn označuje počet mikrojadier v lymfocytoch. Nezávislá premenná BMI je body mass index, sex=0 pre mužov a 1 pre ženy, premenná group má ten istý význam ako v modeli pre sbs. Na úrovni konečného - trvalého poškodenia DNA sa významne prejavil vplyv veku (vekovej skupiny), pohlavia a nadváhy. Tvorba mikrojadier je pravdepodobne iniciovaná oveľa vyššími dávkami poškodzujúcich faktorov. U starších jedincov je v dôsledku dlhodobej expozície vonkajšiemu prostrediu a akumulácii poškodení zvýšená hladina mikrojadier. Model tiež naznačuje zvýšené poškodenie u ľudí s nadváhou. Je to dôsledok úzkeho prepojenia nadváhy s inými ochoreniami na báze oxidačného stresu (napr. ateroskleróza), vplyvom horšieho životného štýlu a schopnosťou toxických látok akumulovať sa v tukovom tkanive. Zvýšená hladina mikrojadier u žien môže

súvisieť s vyšším podielom tukového tkaniva v organizme a odlišným hormonálnych systémom.



Záver: Možno povedať, že oba typy poškodenia sú zviazané s príslušnosťou k vekovej skupine, osoby zo staršej skupiny majú viac poškodení. Pri poškodení sbs záleží aj na prítomnosti génu GSTT1 – v skupine starších znamená zvýšenie poškodenia. Vyššie hodnoty nmn majú ženy a osoby s vyšším BMI Fakt, že sa tieto dva markery poškodenia DNA viažu s odlišnými premennými, je pravdepodobne daný ich rozdielnou biologickou podstatou, citlivosťou detekcie i kvalitatívnou závažnosťou. I na to treba prihliadať pri interpretácii experimentálnych výsledkov.

Použitá literatúra:

Steven N. Austad, Why we age, John Wiley & Sons, Inc., 1997

Adresa: RNDr. Ladislava Wsólová, Slovenská zdravotnícka univerzita, Limbová 12, 83303 Bratislava, ladislava.wsolova@szu.sk

ENVIRONMENTÁLNA KUZNETSOVOVA KRIVKA V EU

Hana ZACH

Abstract
The aim of the article was to develop environmental Kuznets curve and compute the break point in EU using 2005 Environmental Sustainability Index (ESI). The results indicated countries earning more than 14867 USD in purchasing power parity were on track towards environmental sustainability. However, splitting ESI into two parts, feasible EKC shape for higher GDP per capita was reached mainly due to higher social and institutional capacity, technology advances or global stewardship, not by cleaner environment.

Keywords: environmental sustainability index, environmental Kuznets curve, EU

Úvod

Tradične sa argumentuje, že v jadre konceptu udržateľného rozvoja leží rozkol medzi životným prostredím a ekonomickým rozvojom. Predpokladá sa, že vzrastajúce nároky na spoločnosti chrániť životné prostredie len zvyšujú náklady výroby, odvádzajú pozornosť manažmentu od prieraznejších investícií či ich priamo nútia investovať v krajinách s obmedzenou legislatívou životného prostredia. Carter (2000) však poukazuje na pozitívny vplyv ekonomického rastu na kvalitu životného prostredia. Podľa neho neplatí, že environmentálna legislatíva a inovácie technológii smerom k udržateľnému rozvoju brzdia ekonomický rast, nie je pravdou, že je krajina buď bohatá, alebo má čisté životné prostredie, vzájomne vylučujúc mať obe položky naraz. Lekalis a Kousis (2001) dospeli k záveru, že rast HDP na obyvateľa vedie k zlepšeniu životného prostredia. Zmenu k lepšiemu vyvoláva dopyt spotrebiteľov, ktorí vyvíjajú tlak na vládu v snahe získať kvalitné životné prostredie, ktoré je tak chápané ako verejný statok. Vzťahu životné prostredie a ekonomický rozvoj sa venovali aj Jha a Murthy (2003), poukázali na existenciu tzv. environmentálnej Kuznetsovovej krivky (EKC), ktorá práve tento vzťah popisuje. Empiricky odvodený tvar EKC naznačuje, že vyššie HDP na obyvateľa je zrejme cestou k lepšiemu životnému prostrediu. Má tvar invertovaného písmena U, pričom do určitej úrovne príjmu environmentálne znečisťovanie narastá, po prekročení tejto hranice príjmu dochádza k poklesu znečisťovania. Bod zvratu sa vysvetľuje tým, že s rastom HDP na obyvateľa sa uvoľňuje viacej zdrojov na monitoring a vymožiteľnosť práva (Dinda, 2004; Stern, 2003; Urbanová, 2005).

Materiál a metódy

Cieľom práce je definovať environmentálnu Kuznetsovovu krivku v EU priestore s výnimkou Cyprusu a Malty. Skúmame vzťah environmentálnej udržateľnosti a ekonomického rozvoja. Pracujeme s prierezovými údajmi krajín definovanými v 2005 Environmental Sustainability Index (2005). Pracujeme s indexom environmentálnej udržateľnosti (ESI) a podielom HDP na obyvateľa v parite kúpnej sily v USD (HDP), oba údaje sme čerpali z 2005 Environmental Sustainability Index (2005). Esty et al. (2001) popisuje ESI za snahu vyjadriť progres smerom k trvalej udržateľnosti prostredníctvom jedného čísla. Zameriava sa na environmentálnu udržateľnosť, ktorú je možné dosiahnuť pri danom ekonomickom raste. Vyjadruje progres a nie stav, preto okrem „pravých“ ukazovateľov životného prostredia obsahuje aj množstvo technických či sociálnych premenných, ktoré majú zabezpečiť podmienky rýchlej odozvy na daný stav či už v oblasti legislatívy alebo inovácií. Betts (2000) považuje ESI za obdobu HDP, teda snahu vyjadriť jedným číslom realitu ohľadom životného prostredia. ESI 2005

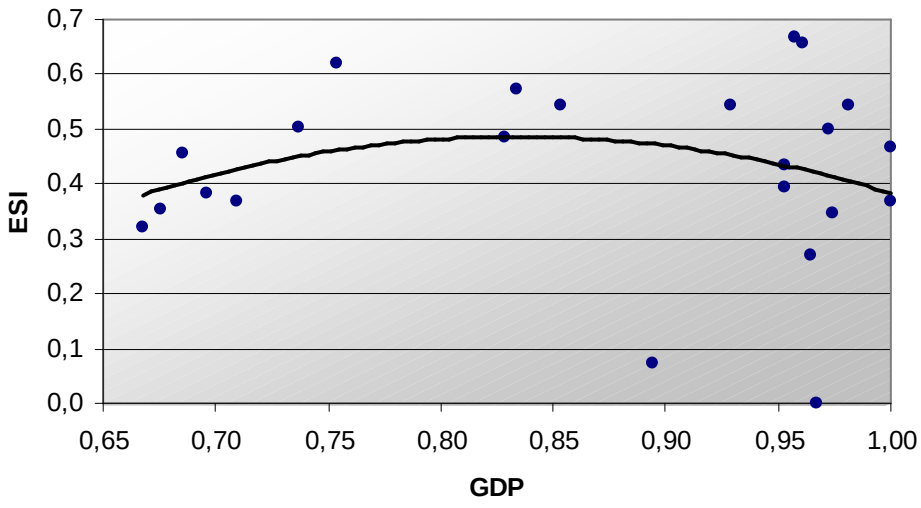
agreguje 21 indikátorov životného prostredia zostavených z pôvodných 76 individuálnych premenných do piatich hlavných komponentov (environmentálny systém, redukovanie environmentálnych záťaží, redukovanie ľudskej zraniteľnosti, sociálna a inštitucionálna kapacita, globálna spolupráca). ESI hodnota ako aj hodnoty piatich komponentov kolíšu medzi 0 a 100, pričom vyššia hodnota znamená lepšiu pozíciu pre environmentálnu udržateľnosť.

Z dôvodu rôznych mier skúmaných znakov transformujeme pôvodné štatistické znaky do funkcií príslušnosti fuzzy množín A(X) (Nech U je množina a L = (L, $\wedge$, $\vee$, 1, 0) zväz. Funkciu A definovanú na univerzu U, t.j. A: U $\rightarrow$ L, nazývame fuzzy množinou, presnejšie funkciou príslušnosti množiny A. L definujeme ako L = <0, 1>). Touto úpravou sa vyhýbame prisudzovaniu vyššej dôležitosti znakom pohybujúcim sa vo vyšších hodnotách. Vzhľadom k pôvodnému vykresleniu EKC meníme smer ESI a jeho piatich komponentov. Úpravou 1-A(X) získame požadovaný tvar, pričom vyššia hodnota funkcie príslušnosti fuzzy množiny teraz znamená vyššiu úroveň environmentálnej degradácie (t.j. nižšia hodnota je žiadúca).

Výsledky a diskusia

Vykreslením závislosti environmentálnej degradácie od HDP vo funkciách príslušnosti fuzzy množín dvadsiatich troch krajín EU získavame obrázok 1. Preložením získaných bodov vhodným trendom (parabola) a postavením jej prvej derivácie rovnej nule získavame bod zlomu. Konštatujeme, že do úrovne príjmu 14867 USD v parite kúpnej sily dochádza k nárastu environmentálnej degradácie v krajinách EU, po dosiahnutí tohto príjmu však zaznamenávame jej následný pokles.

Obrázok 1 Environmentálna Kuznetsovova krivka v krajinách EU

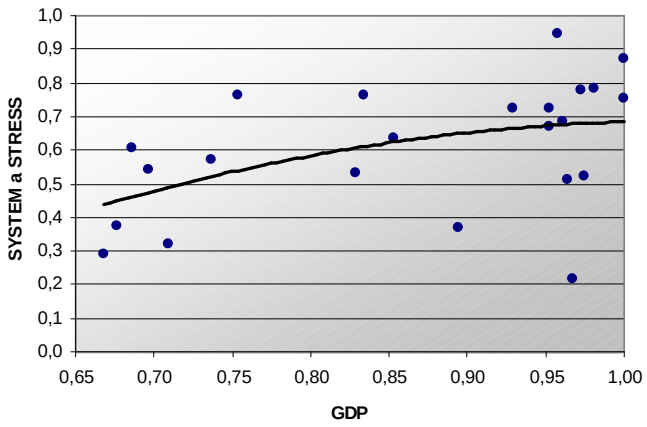


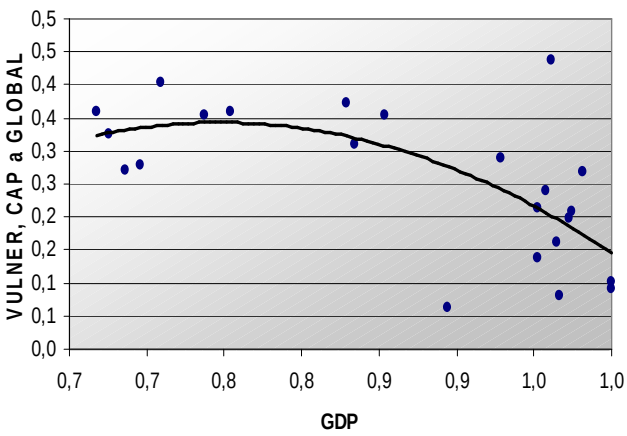
Medzi krajiny patriace do skupiny s príjmom do 14867 USD v parite kúpnej sily, t.j. tie kde dochádza k nárastu environmentálnej degradácie, zaraďujeme Lotyšsko, Litvu, Portugalsko, Slovinsko, Slovensko, Maďarsko, Českú republiku a Španielsko. Krajiny s príjmom nad 14867 USD v parite kúpnej sily, t.j. tie kde dochádza k poklesu úrovne environmentálnej degradácie zaraďujeme Grécko, Švédsko, Taliansko, Francúzsko, Nemecko, Belgicko, Poľsko, Rakúsko, Fínsko, Luxemburg, Írsko, Veľkú Britániu, Dánsko a Holandsko (obr. 2).

Konštatovaním záveru zo získanej skutočnosti by sme sa ochudobňovali o nasledovné zaujímavé informácie. Analyzujeme závislosť HDP v parite kúpnej sily v USD a dvomi skupinami komponentov ESI, do prvej skupiny zaraďujeme komponenty environmentálny systém (SYSTEM) a redukovanie environmentálnych záťaží (STRESS), čo sú komponenty priamo sa dotýkajúce stavu životného prostredia, druhú skupinu tvoria komponenty redukovanie ľudskej zraniteľnosti (VULNER), sociálna a inštitucionálna kapacita (CAP) a globálna spolupráca (GLOBAL), čo sú komponenty vyjadrujúce pripravenosť implementovať stratégie rozvoja životného prostredia.

Obrázok 2 Rozdelenie krajín EU podľa bodu zlomu EKC

Obrázok 3 a 4 HDP/obyv. v PPP USD na SYSTEM, STRESS a VULNER, CAP, GLOBAL





Ak by sme sa zaoberali iba skúmaním stavu životného prostredia v závislosti na HDP, zistili by sme, že s rastom HDP rastie aj zhoršovanie stavu životného prostredia a bod zlomu je až pri hodnote 50867 USD na obyvateľa v parite kúpnej sily (obr. 3). Ak by sme sa zaoberali iba skúmaním pripravenosti implementovať stratégie rozvoja životného prostredia, zistili by sme, že s rastom HDP rastie aj pripravenosť krajín s bodom zlomu pri hodnote 8767 USD na obyvateľa v parite kúpnej sily (obr. 4). Výpočtami EKC v rámci EU sa potvrdilo to o čom písal už Arnold (2002), ktorý tvrdil, že kvalita životného prostredia nezávisí len od úrovne ekonomického rozvoja jednotlivých krajín, hoci ide o jeden z dôležitých determinantov. Kvalita životného prostredia sa dá podľa neho zvýšiť napr. technickým pokrokom, ktorý zväčša environmentálne následky nemá. Napríklad Stern (2003) písal, že pokles znečistenia s rastom HDP na obyvateľa nesúvisí priamo, t.j. bohatšie krajiny nemajú čistejšie životné prostredie len preto že sú majetnejšie. Ekonomickou silou však vedia investovať do technológií a takých opatrení, do ktorých menej majetné krajiny investovať nemôžu. Ďalej sú to štrukturálne zmeny smerom k väčšej informovanosti, vymožiteľnosť práva, rôzne environmentálne regulácie a vyššie výdavky smerom k ochrane životného prostredia, čo všetko vyúsťuje do postupného znižovania environmentálnej degradácie. Takisto Urbanová (2005) uviedla, že dôvodom čistejšieho životného prostredia sú do veľkej miery zákony, regulácie a normy, ktoré vzišli z demokratického usporiadania. Podľa nej je veľmi dôležitá vymožiteľnosť práva, pretože štáty s nízkou vymáhateľnosťou neposkytujú dostatočnú ochranu pred znečisťovateľmi a následným znečistením, nakoľko náklady spojené s rizikom postihu v týchto štátoch sú nižšie ako výnosy z takéhoto chovania.

Záver

Na základe vypracovaných analýz sme vykreslili environmentálnu Kuznetsovovu krivku (EKC) dvadsiatich troch krajín EU s bodom zlomu pri 14867 USD v parite kúpnej sily. Rozložením indexu environmentálnej udržateľnosti na dve časti však zisťujeme, že pre krajiny EU je jednoduchšie tvoriť stratégie pripravenosti a budovať inštitúcie zabezpečujúce ochranu životného prostredia ako zabezpečiť samotný stav životného prostredia. Ak by sme hodnotili len stav životného prostredia (SYSTEM a STRESS), konštatovali by sme, že ani jedna krajina EU bod zlomu na EKC zatiaľ nedosiahla, t.j. s rastom HDP rastie aj úroveň environmentálnej degradácie. Pokles environmentálnej degradácie znázornený pomocou EKC pri vyššej hodnote HDP na obavateľa je tak predovšetkým dosiahnutý pripravenosťou krajín, technológiou a cezhraničnou spoluprácou.

Literatúra

1. ARNOLD, R. 2002. Development, Environmental Health, and English Cooking. In: Journal of Environmental Engineering, Vol. 128, Issue 7, p. 571-574.
2. BETTS, K.S. 2000. New sustainability benchmark takes aim at the GDP. In: Environmental Science & Technology, Vol. 34, Issue 7, p. 165A.
3. CARTER, P. 2000. World Watch: Living Green and the Bottom Line. In: Mother Earth News, Issue 180, s. 14-16.
4. DINDA, S. 2004. Environmental Kuznets Curve Hypothesis: A Survey. In: Ecological Economics, Vol. 49, Issue 4, s. 431-456.
5. ESTY, D. - SAMUEL-JOHNSON, K. 2001. Environment by number. In: World Link, Vol. 14, Issue 1, p. 23-26.
6. ESTY, D. C. - LEVY, M. - SREBOTNJAK, T. - DE SHERBININ, A. 2005 Environmental Sustainability Index: Benchmarking National Environmental Stewardship. New Haven: Yale Center for Environmental Law and Policy, 2005.

Available on the website http://www.ciesin.columbia.edu/indicators/ESI [2005-11-05]
7. JHA, R. - MURTHY, K.V.B. 2003. An inverse global environmental Kuznets curve. In: Journal of Comparative Economics, Vol. 31, Issue 2, s. 352-369.
8. LEKAKIS, J.N. - KOUSIS, M. 2001. Demand for and supply of environmental quality in the environmental Kuznets curve hypothesis. In: Applied Economics Letters, Vol. 8, Issue 3, s. 169-173.
9. STERN, D. 2003. The Environmental Kuznets Curve. International Society for Ecological Economics. Rensselaer Polytechnic Institute. Dostupné na internete http://www.ecoeco.org/publica/encyc_entries/Stern.doc; [2005-05-15]
10. URBANOVÁ, T. 2005. Rizika státního managementu ochrany životního prostředí. In: Environmentálna politika v SR. Transparentnosť a environmentálna politika. Zborník. Bratislava: Konzervatívny inštitút Milana Rastislava Štefánika. 2005. ISBN 80-89121-07-1, s. 68-85

Adresa autorky
Ing. Hana Zach, SPU FEM KŠOV Nitra, hana.zach@fem.uniag.sk

Využití stromů v analýze dat

Marta Žambochová

Abstract:

We know a way of knowledge representation in a form of trees from many spheres such as computer science (data structure), biology (classification), psychology (decision theory), social science (organization structure) and many other fields. This paper discuss the implementation of decision trees in statistical data analysis. It describes two sorts of decision trees - classification and regression trees and their possibility to use in a regression analysis, pattern recognition problem, a cluster analysis and a time series analysis. Using decision trees has many advantages, such as its fast speed, high accuracy, simplicity of interpretation, opportunity of processing both quantitative and qualitative information.

Úvod

Využití stromových struktur je známo z různých oblastí běžného života. Jedná se o různé organizační „pavouky" vystihující co nejpřesněji hierarchické uspořádání sledovaného objektu. Zobrazujeme takto organizační strukturu nějaké instituce, hierarchické uspořádání adresářů v počítači, klasifikace biologických objektů a mnohé další.

Stromy jsou spojité acyklické grafy. Skládají se z uzlů (vrcholů) a hran (spojnic vrcholů). Kořenový strom je strom, v němž je jeden z vrcholů vybrán jako hlavní (počáteční) vrchol. Tomuto vrcholu se říká kořen. Stupněm vrcholu nazýváme počet hran incidentních s daným vrcholem. Vrcholy stupně 1 (koncové vrcholy) nazýváme listy.

Většinou se kořenový strom zobrazuje ve směru shora dolů (resp. z leva doprava) v pořadí kořen – nekoncové (vnitřní, neterminální)) uzly – listy.

Velmi rozšířenou skupinou stromů, kterých se využívá v datových modelech, jsou různé typy rozhodovacích stromů. Rozhodovací stromy jsou struktury, které rekurzivně rozdělují zkoumaná data dle určitých rozhodovacích kritérií.

Údaje týkající se vysvětlované proměnné jsou obsaženy v listech. Kořen koresponduje s celým vzorkem zkoumaných dat, vnitřní uzly korespondují s podmnožinou vzorku dat z uzlu přímo nad ním a reprezentují vždy jedno kritérium pro další dělení dat do skupin. Do stromového datového modelu ještě připojujeme popis jednotlivých uzlů a hran (větví při štěpení daného vzorku dat) pomocí údajů o vysvětlované a vysvětlujících proměnných.

Vytváření stromu

Rozhodovací strom se vytváří rekurzivně dělením prostoru hodnot prediktorů (vysvětlující, nezávislé proměnné).

Máme-li strom s jedním listem, hledáme otázku (podmínku větvení), která nejlépe rozděluje prostor zkoumaných dat do podmnožin, tj. maximalizuje kritérium kvality dělení (tzv. splitting criterium).

Takto nám vznikne strom s více listy. Nyní pro každý nový list hledáme otázku, která množinu prediktorů náležící tomuto listu co nejlépe dělí do podmnožin.

Proces dělení se zastaví, pokud bude splněno kritérium pro zastavení (tzv. stopping rule). Omezení obsažená v kritériu pro zastavení mohou být např. „hloubka“ stromu, počet listů stromu, stupeň homogennosti množin dat v listech, …

Jednotlivé algoritmy vytváření rozhodovacích stromů se liší následnými charakteristikami:

- pravidlo dělení
- kritérum pro zastavení
- typ podmínek větvení
  - multivariantní (testuje se několik prediktorů)
  - univariantní (v daném kroku se testuje pouze jeden z prediktorů)
- způsob větvení
  - binární (každý z uzlů, kromě listů, se dělí na dva následníky)
  - k-ární (některý z uzlů se dělí na více než dvě části)
- typ výsledného stromu, popis obsahu listů
  - klasifikační stromy (v každém listu je přiřazení třídy)
  - regresní stromy (v každém listu je přiřazení konstanty – odhad hodnoty závislé proměnné)
- typ prediktorů
  - kategoriální
  - ordinální

Použití rozhodovacích stromů v analýze dat

Jedním z typů úloh, kde je možno k řešení využít rozhodovacích stromů je klasifikace. Tedy úlohy, ve kterých hledáme model, kde závislou proměnnou zařazujeme na základě hodnot nezávisle proměnných do určitých tříd. K tomuto účelu se využívají klasifikační stromy.

Oblastí tohoto typu je diagnostika. Příkladem takovéto praktické úlohy je potřeba banky rozdělit své zákazníky, zájemce o úvěr, na důvěryhodné a nedůvěryhodné, nebo potřeba firmy pojmenovat charakteristiky jednotlivých typů svých zákazníků pro určení směru dalšího možného vývoje firmy.

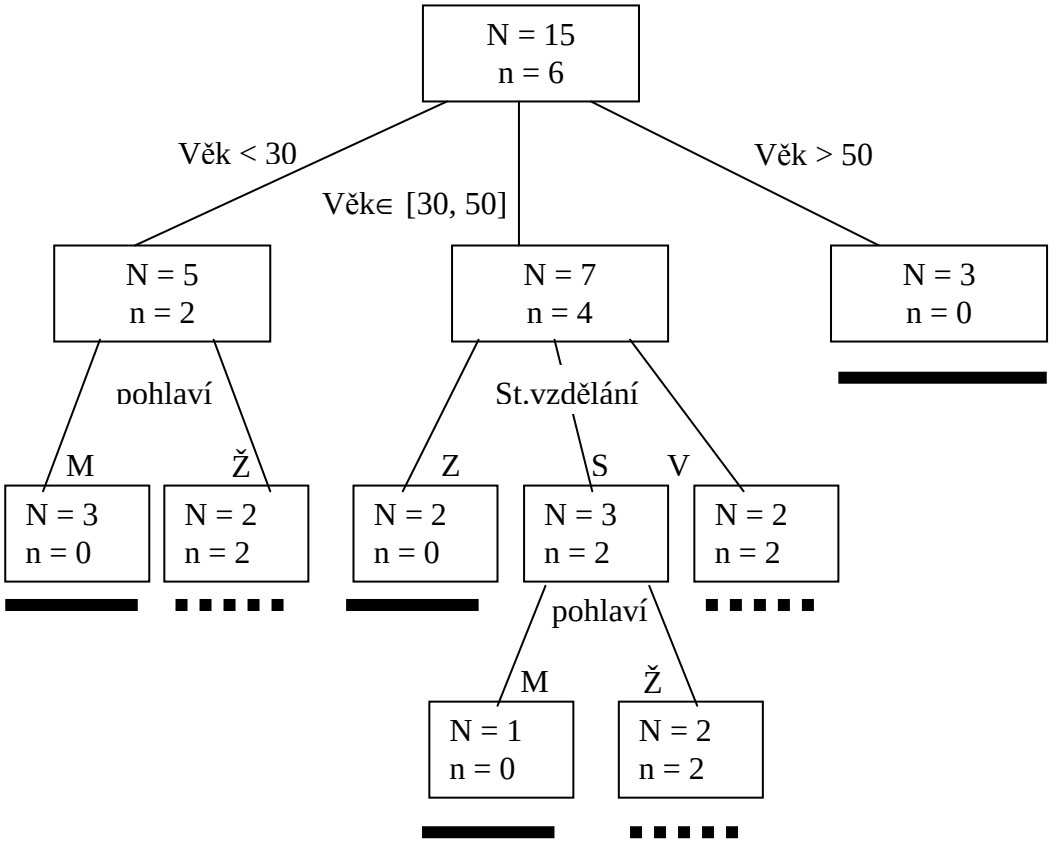
S potřebou klasifikace (zařazení do tříd) se můžeme setkat i mnoha dalších oblastech.

Příklad 1.

Následující příklad ukazuje využití rozhodovacích stromů při klasifikaci. Máme za úkol respondenty rozdělit do dvou tříd – ty kteří mají zájem o námi sledovanou akci a ty, kteří o akci zájem nemají. Máme k dispozici následující údaje o jednotlivých respondentech.

Respondent	Věk (X_1)	Pohlaví (X_2)	Stupeň vzdělání (X_3)	Zájem (Y)
1	18	M	S	N
2	31	Ž	Z	N
3	50	Ž	V	A
4	45	M	S	N
5	20	M	V	N
6	36	M	V	A
7	66	Ž	S	N
8	24	M	Z	N
9	35	Ž	S	A

10	21	Ž	S	A
11	48	M	Z	N
12	32	Ž	S	A
13	79	M	V	N
14	58	M	S	N
15	19	Ž	Z	A



N … počet respondentů obsažených v uzlu celkem
a … počet respondentů obsažených v uzlu, kteří mají zájem o sledovanou akci
▬ … list obsahující pouze respondenty nemající zájem … 1.třída
▪▪▪▪ … list obsahující pouze respondenty mající zájem … 2. třída

---

Dalším typem úloh analýzy dat, kde můžeme využít rozhodovacích stromů je odhadování hodnot vysvětlované proměnné. Odhadovaná proměnná může být jak kategoriální, tak spojitá. V prvním případě použijeme k vytváření modelu stromy klasifikační, v druhém pak stromy regresní.

V praxi se s daným typem úloh můžeme setkat při marketingovém rozhodování. Například tehdy, pokud potřebujeme vytipovat potenciální zákazníky, u kterých je největší naděje na kladnou odezvu (např. koupi výrobku). Z celé velké databáze potencionálních zákazníků oslovíme nabídkou náhodně a reprezentativně vybranou skupinu. Výsledky nabídky analyticky zpracujeme a vytipujeme charakteristické vlastnosti osob, které reagovaly

pozitivně. Osoby s těmito charakteristickými vlastnostmi poté cíleně oslovíme v reklamní kampani.

Příklad 2.

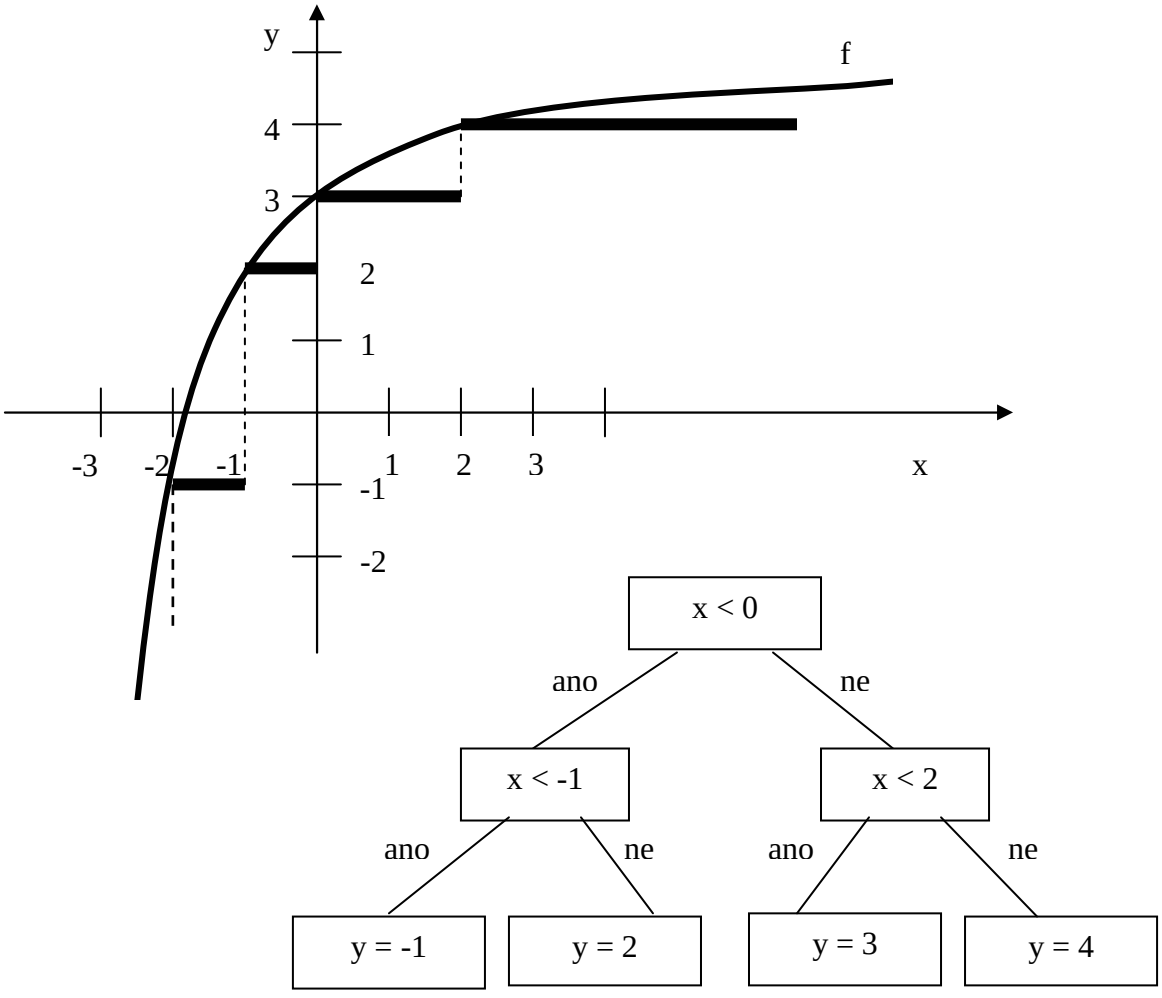
Pokud bychom na základě výsledků z předchozí ho příkladu chtěli vyvodit závěry, mohli bychom předpokládat, že zájem budou mít následující typy osob:

1. ženy mladší 30 let
2. ženy středoškolského vzdělání věku mezi 30 a 50 lety
3. osoby vysokoškolsky vzdělané ve věku mezi 30 a 50 lety

Regresní stromy jsou alternativou klasické regresní analýzy v případech, kdy nám postačuje odhad regresní křivky po částech konstantní „schodovitou“ funkcí. Přínosem je názornost a lehká interpretace.

Příklad 3.

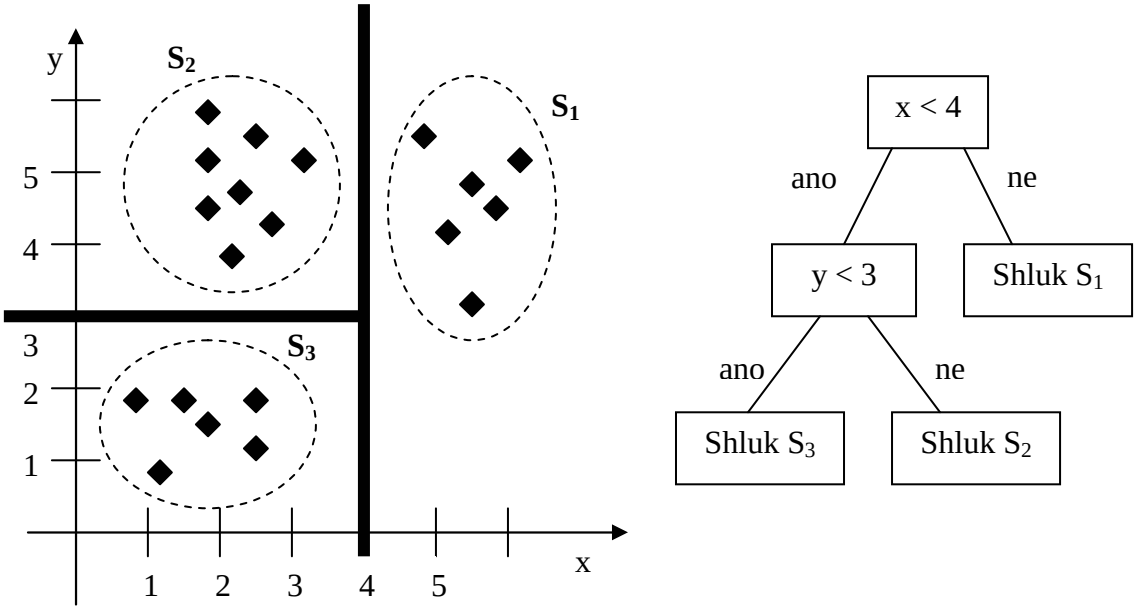
Regresní strom na následujícím obrázku je po částech konstantní aproximací regresní funkce f.



Rozhodovací stromy jsou použitelné i pro úlohy typu segmentace (shlukování). Cílem této skupiny úloh je rozdělení datového souboru dat do skupin (shluků), přičemž objekty uvnitř jednotlivých shluků jsou si co nejvíce podobny, zatímco objekty z různých shluků si jsou podobny jen minimálně.

V praxi se s daným typem úlohy můžeme setkat například při potřebě rozčlenit nesourodý trh na několik skupin - shluků. Spotřebitelé zařazeni do jednoho shluku vykazují podobné chování a preference, a proto je snazší pro ně optimalizovat marketingový a komunikační mix.

Příklad 4.



V neposlední řadě se rozhodovací stromy dají využít i při predikci v časových řadách. Tedy v případě, že máme předpovědět nějaké (kvantitativní nebo kvalitativní) charakteristiky v nějakém budoucím časovém okamžiku na základě hodnot všech těchto charakteristik v minulosti.

Literatura

1. Hebák, P., Vícerozměrné statistické metody. Praha Informatorium 2004.
2. Antoch J., Klasifikace a regresní stromy. Sborník ROBUST 88
3. Keprta, S. , Nebinární klasifikační stromy. Sborník ROBUST 94
4. Savický, P., Klaschka, J., a Antoch J., Optimální klasifikační stromy. Sborník ROBUST 2000
5. Spousta, J., Jak poznat toho pravého čili rozhodování na stromech. http://www.spss.cz/files/programy/at_text.pdf
6. Berikov, V., Litvinenko, A., Methods for statistical data analysis with decision trees http://www.math.nsc.ru/AP/datamine/eng/decisiontree.htm

Adresa

RNDr. Marta Žambochová
Fakulta sociálně ekonomická, Univerzita J.E. Purkyně, Ústí nad Labem
Katedra matematiky a statistiky
zambochova@fse.ujep.cz

Data mining a metóda GUHA

Ivan Žežula

Abstract: *We propose several improvements of GUHA data mining method, namely a class of better implicational quantifiers and some better association quantifiers.*

Princíp

Metóda GUHA (General Unary Hypotheses Automaton), známa aj pod názvom 4ft-Miner, je dnes už klasickou metódou data miningu, zameranou najmä na získavanie informácií z diskrétnych dát. V prípade spojitých dát sa najprv robí ich diskretizácia. Princíp metódy je nasledujúci:

1. Vstupné dáta sa predpokladajú v tvare matice, kde riadky tvoria pozorovania na jednotlivých objektoch a stĺpce sú sledované premenné, t.j. obvyklá databázová štruktúra. Ak je premenná *Var* dichotomická, bez ujmy na všeobecnosti ju pokladáme za 0-1 a *Not Var* označuje jej binárnu negáciu. V prípade multinomickej premennej *Var:Cat* označuje binárnu premennú, ktorá je indikátorom príslušnej kategórie *Cat* premennej *Var*; *Not Var:Cat* je jej binárna negácia.
2. Automaticky sa generujú všetky možné **elementárne konjunkcie**

   $$Lit\_1 \,\&\, Lit\_2 \,\&\, ... \,\&\, Lit\_m,$$

   kde *Lit_j* je **literál**, čiže výraz typu *Var, Not Var, Var:Cat* alebo *Not Var:Cat.* Všetky premenné sa takto charakterizujú pomocou binárnych premenných.
3. Elementárne konjunkcie sa potom používajú na generovanie všetkých možných hodnôt **hypotéz**, ktorých typy si zvolí používateľ. Všeobecný tvar hypotézy je

   $$Ant \square Succ \mid Cond,$$

   kde Ant (antecedent), Succ (succedent) a Cond (condition) sú elementárne konjunkcie a □ je binárny **kvantifikátor**. Najčastejšie používané typy kvantifikátorov sú:
   a. implikačný →
   b. asociačný ≈
   c. korelačný #
4. V prípade, že vstupné dáta nemajú chýbajúce údaje, každý kvantifikátor vygeneruje kontingenčnú tabuľku

	Succ	**Not Succ**	**Σ**
Ant	a	b	r
Not Ant	c	d	s
Σ	k	l	n

   na základe ktorej sa vyhodnotí jeho pravdivosť. Kvantifikátor je teda binárna funkcia

F(a,b,c,d), hodnotiaca pravdivosť hypotézy *Ant* □ *Succ* za podmienky *Cond* (vymedzujúcej podmnožinu dátovej matice) v dátach. Väčšina používaných kvantifikátorov je založených na štatistických testoch.

Implikačné kvantifikátory

Veľmi často používaný kvantifikátor je **fundovaná implikácia**: má parametre *base* a *cp*. Implikácia je pravdivá (splnená v dátach), ak

$$a \geq base \quad \& \quad a/r \geq cp.$$

Je zrejmé, že pre *base=0* a *cp=1* ide o obyčajnú implikáciu. Podiel a/r predstavuje podmienenú „pravdepodobnosť“ $f(Succ|Ant) = \frac{f(Ant \cap Succ)}{f(Ant)}$. Parameter *base* má ochrannú funkciu – zaručuje, že implikácia sa bude týkať nezanedbateľného počtu prípadov.

Tento operátor neberie do úvahy logickú ekvivalenciu *(Ant → Succ)* ≡ *(Not Succ → Not Ant)*. Mohli by sme teda analogicky uvažovať podiel $f(\neg Ant|\neg Succ) = \frac{f(\neg Ant \cap \neg Succ)}{f(\neg Succ)}$. Takto môžeme vytvoriť kvantifikátor **fundovaná obmenená implikácia**, ktorá bude pravdivá v prípade, že

$$d \geq base \quad \& \quad d/l \geq cp.$$

Tento samotný nemá žiadny praktický význam, keďže v procese generovania všetkých možných antecedentov a succedentov sú pre každú dvojicu vygenerované niekedy aj ich negácie. Môžeme však vytvoriť naraz ich vážený priemer: **fundovaná vážená implikácia** s váhou $\alpha \in (0;1)$ je pravdivá v prípade, keď

$$\alpha\, a+(1-\alpha)d \geq base \quad \& \quad \alpha\, a/r+(1-\alpha)d/l \geq cp.$$

Špeciálny prípad je α=1/2, môžeme ho nazvať **priemernou fundovanou implikáciou**:

$$(a+d)/2 \geq base \quad \& \quad a/r+d/l \geq 2cp.$$

Asociačné a korelačné kvantifikátory

Rozdiel medzi asociačnými a korelačnými kvantifikátormi je zrejme hlavne v tom, že korelačné pracujú priamo so spojitými veličina, t.j. nerobí sa diskretizácia. Cieľom oboch však je zistiť mieru závislosti resp. nájsť v dátach významne závislé veličiny.

Asociačné kvantifikátory hľadajú závislosti medzi elementárnymi konjunkciami. Pritom sa zbytočne generujú aj triviálne závislosti medzi binárnymi premennými príslušnými k jednej kategoriálnej premennej. Na druhej strane korelačné kvantifikátory používajú ako vstupné veličiny iba jednotlivé spojité premenné. Medzi tým je v doteraz používanej metodike územie nikoho.

V záujme lepšieho využitia informácií v dátach by bolo dobré pripustiť väčšiu pestrosť typov dát. To by vyžadovalo, aby v prvej fáze spracovania každá premenná pri sebe niesla aj

informáciu o svojom type. Navrhujeme používať tieto typy:

- nominálne
  - binomické
  - multinomické
- ordinálne
- kvantitatívne
  - diskrétne
    - so spojitým generátorom
    - s nespojitým generátorom
  - spojité
    - normálne
    - nenormálne

Táto klasifikácia, ktorej zadanie by systém vyžadoval, by mala umožniť automatický výber vhodnej metódy výpočtu asociačného alebo korelačného kvantifikátora.

V prípade multinomických premenných by sa nevytvárala kontingenčná tabuľka 2×2, ale všeobecná $r \times c$. Z nej by sa potom dali definovať kvantifikátory založené na známych mierach asociácie, ako sú (upravený) kontingenčný koeficient, Cramerovo V, koeficienty λ, či koeficienty neurčitosti. Elementárne konjunkcie by bolo možné naďalej používať do podmienky *Cond* na zúženie množiny pozorovaní.

Napr. **Cramerov kvantifikátor** pre dve multinomické premenné s kategóriami A_i a B_j je založený na tabuľke

$A \backslash B$	B_1	B_2	. . .	B_c	Σ
A_1	n_{11}	n_{12}	. . .	n_{1c}	n_{10}
A_2	n_{21}	n_{22}	. . .	n_{2c}	n_{20}
. . .	. . .	. . .	. . .	. . .	. . .
A_r	n_{r1}	n_{r2}	. . .	n_{rc}	n_{r0}
Σ	n_{01}	n_{02}	. . .	n_{0c}	n

Jeho parametre sú *base* a k_α. Kvantifikátor je pravdivý v prípade, že

$$n \geq base \;\; \& \;\; \sqrt{\frac{\chi^2}{n \cdot \min(r-1, c-1)}} \geq k_\alpha ,$$

kde χ^2 je štandardná χ^2-štatistika a k_α vhodne zvolená konštanta.

Kontingenčné tabuľky typu $r \times c$ sa dajú použiť na konštrukciu asociačných a korelačných kvantifikátorov aj v prípade ordinálnych premenných, v tomto prípade založených na koeficiente γ, Kendallovom a Stuartovom τ, Somersovom *d*, či Spearmanovom korelačnom koeficiente.

V prípade kvantitatívnych diskrétnych premenných je dôležité rozlíšiť, či proces, ktorý ich generuje, má spojitý alebo diskrétny charakter. Niekedy totiž pozorujeme diskrétne veličiny, či dokonca iba ordinálne, ale v pozadí sú nepozorované či nepozorovateľné spojité (latentné)

premenné. V takom prípade je vhodné používať polychorické korelácie. V opačnom prípade môžeme voliť medzi Pearsonovým (najmä pri vyššom počte rôznych hodnôt), Spearmanovým a Kendallovým korelačným koeficientom. Posledné dva sa odporúčajú aj pre nenormálne rozdelené spojité dáta, kým Pearsonov koeficient je ideálny v prípade normálnych premenných.

Dôležitý je aj prípad, keď jedna premenná je spojitá a druhá nominálna. Vo takom prípade sa dá použiť koeficient nelineárnej korelácie η, pričom na mieste nominálnej premennej môže byť aj elementárna konjunkcia. V prípade jednej ordinálnej a jednej spojitej premennej môžeme použiť polyseriálny korelačný koeficient. Konštrukcia príslušných kvantifikátorov je zrejmá.

Záver

V posledných rokoch sa na poli korelačnej a asociačnej analýzy odohráva búrlivý vývoj, ktorý by mali zachytiť aj moderné data miningové metódy. Keďže jeden vývoj sa odohráva najmä v oblasti sociológie a psychológie a druhý predovšetkým v oblasti ekonómie, komunikácia trocha zaostáva. Ukazuje sa tiež, že vo väčšej miere ako doteraz je potrebné využívať informácie o dátových typoch, ktoré zohľadňujú aj proces vzniku hodnôt príslušných premenných. To umožňuje konštruovať jemnejšie metódy, citlivejšie reagujúce na hľadané javy.

Doc. RNDr. Ivan Žežula,CSc.
UMV PF UPJŠ,
Jesenná 5,
041 54 Košice
zezula@kosice.upjs.sk

Táto práca bola podporená grantom VEGA MŠ SR č. 1/0385/03.

Statistické výpočty 2005

Jiří Žváček

Práce statistika zahrnuje

- sběr a úpravu dat,
- interaktivní zpracování dat a
- publikování výsledků.

Potřebuje tedy buď velmi komplexní nástroj nebo několik jednoúčelových. Komplexně se snaží řešit celou problematiku **statistické pakety**.

Všechny úlohy však nemůže pokrýt žádný jednotlivý program, takže znalost široké plejády programů je podmínkou statistikovy práce a je třeba využívat a znát i software pocházející primárně z jiných oblastí. Vzniká tedy **statistické výpočetní prostředí**, které si ovšem každý statistik podle svých potřeb a znalostí ale zejména možností přizpůsobuje.

Co používají statistikové v Čechách ilustruje **anketa o statistickém softwaru** (www.stahroun.me.cz/statisticka_anketa/anketa-formular.php) na stránce České statistické společnosti.

1. Novinky obecně

Vývoj softwaru v oblasti v oblasti statistického výpočetního prostředí býval pro softwarové specialisty idylkou ve srovnání s některými jinými oblastmi. Statistika nemá až na výjimky **open source** s krátkou a nepravidelnou periodicitou a používané programy jsou pro svou minimální rozšířenost ušetřeny specializovaných virů (rozšířen je spíše crack).

Tato idylka však končí a novinky se stávají nejenom častější, ale je jich i mnohem více. Jsou k tomu zejména tyto důvody:

Internetová komunikace
: Produkty a služby jsou okamžitě celosvětově dostupné. Ve většině oborů se velmi zrychlila výměna informací a v důsledku toho i frekvence inovací.
I publikování se pomalu přesouvá na web a informace a novinky se už dozvíme mnohem rychleji než z papírových publikací.

Zapojení uživatelů
: Vznikají zájmové a diskusní skupiny a s nimi spojené inovace a nové služby, protože producenti a distributoři software tento proces aktivně podporují a využívají.

Technický pokrok oblasti hardware
: V oblasti **hardware** jsou šlágrem letošního roku zejména **64 bitové** a **vícejádrové** procesory i pro PC počítače. Veškerý numerický a grafický software se bude muset přepsat pro **paralelní výpočty** (letos to stihla MATHEMATICA 5.2, MATLAB SP 2 a STATA).
Pro statistiky to znamená i mnohem delší periodicitu generátorů pseudonáhodných čísel, velké zrychlení grafiky a výpočtů a obrovské databáze.

Pokrok v software
: I statistický **software** se přizpůsobuje tendencím vznikajícím v nejpokročilejších produktech. Nové verze produktů sice stále mají zhruba roční periodicitu, ale objevuje

se již i **SP** (Service Pack) nebo **update**, který nahrazuje inovace za desetinnou tečkou a zdarma opravuje a rozšiřuje možnosti (například MATLAB 14 už má 3 SP).
Revoluci připravují již na příští rok Google a Microsoft - práci se softwarem na internetu (Windows Live, Office Live, Open Office, viz **Gates** (www.novinky.cz/internet/68743-gates--office-a-windows-nabidneme-zive-pres-internet.html)).

Koncentrace a globalizace

Nadnárodní **statistické softwarové firmy** jsou dalším důvodem ke zrychlování tempa novinek.
Vznikl **komerční statistický softwarový podnik**, který má všechny cnosti (i necnosti) průmyslového a obchodního podniku. Tyto firmy už nevedou statističtí nadšenci a chovají se stejně jako jiné monopoly. Požírají a vytlačují slabší konkurenty a používají obvyklé triky jako dumping, reklamu, neustálé drobné inovace, drobné dárečky, akční slevy obaly atd.
Pro ekonoma je radostné pozorovat, jak se stírá rozdíl mezi prodejem statistického softwaru a dámského prádla se svými každoročními kolekcemi, kde rozhoduje grafika.
Pro inovace je ale dobré, že velké firmy mají prostředky na rychlé inovace, i když cílem je zisk a musí si některé inovace šetřit na příště.

Expanze služeb

Okruh činnosti softwarových firem rozšiřuje o další služby (placené i zdarma) jako je technická podpora, průběžné inovace atd. Firmy expandují a provádějí činnosti, které dříve realizovaly pouze vysoké školy:

Školení a semináře

placené i zdarma (prezentace). Firmy si troufají i na **konference**, mezinárodní jsou třeba každoroční **SPSS directions** (www.spss.com/spssdirections/) či **SPPS Data Mining Day** (www.spss.com/dataminingday/)

Výuka

a to nejenom práce s paketem ale i statistických metod a to i pokročilých či specializovaných ve vlastních učebnách i u zákazníka (dokonce na vysokých školách).

Publikování

vydávají **časopisy** o nových metodách a aplikacích, **knihy** o programech i obecně statistické **vizuální prezentace** atd.

Objevují se nové služby, zejména založení na internetu, kterými se snaží tyto firmy připoutat potenciální zákazníky. Příkladem může být **Webinar (**je zkratka z Web-based seminář), interaktivní seminář založený na internetu (viz **Wikipedia** (www.webopedia.com/TERM/W/Webinar.htmlv)), který se používá k výuce a dotazům na produkt. Viz třeba přehled na **JMP** (www.jmp.com/news/webinars.shtml)
Prezenčně bývají tyto prezentace v určitém čase zdarma a je možno se interaktivně dotazovat, záznam bývá často placený.

Obdobné firmy nemusí nutně softwarové programy vyvíjet. Zejména je to patrné u **lokálních zastoupení**, které využívají své relativní samostatnosti k obdobné činnosti v národních jazycích.
Nejpokročilejší je tento proces překvapivě v Rusku, kde místní pobočka **Softline** (www.softline.ru/) (původně paket STATISTICA) spravuje mnoho matematických paketů a dokonce vytvořila výukové servery **exponenta.ru** (www.exponenta.ru/) pro matematické a statistické pakety a výuku a **it-university.ru** (www.it-university.ru/) ve spolupráci s univerzitou pro výuku informačních technologií.

Obdobně však postupují i jiné firmy, výuku ke svému produktu nabízí třeba **Navrcholu.cz** (navrcholu.cz/)

Aplikace

Zejména u obecnějších systémů lze vytvářet specializované aplikace, které na bázi daného systému řeší konkrétní úlohu.
Publikování aplikací, které mohou dodávat i zákazníci podstatně zvyšuje četnost novinek (zejména MATHEMATICA, MATLAB).

Vzhledem k množství novinek je třeba je sledovat prakticky denně. Znamená to buď stránky denně prohlížet, nebo použít sledovací programy či nechat si posílat **internetový časopis** nebo **emailové novinky**. I do této oblasti proniklo **RSS**, jednoduché sdílení zpráv, které umožňuje automatické sledování novinek na stránkách statistických softwarových firem (např. MATLAB).

2. Novinky v oblasti paketů

V oblasti statistických metod pakety vyplňují mezery, provádějí drobné úpravy v ovládání a zlepšují grafiku.

Za zmínku stojí dva jevy

Internetový server

který umožňuje **oprávněným osobám** práci s paketem a připravenými aplikacemi v prostředí internetu (tedy třeba doma).
Takový server mají všechny významnější pakety (SPSS, S+, SAS, STATGRAPHICS).

Dynamická internetová grafika

I v paketech se stále více využívá interaktivní úprava a modifikace grafu.

2.1. Nejvýznamnější statistické pakety

Několik paketů dominuje na světovém trhu zejména díky tradici, které si mohou dovolit rozsáhlou komplexní podporu a v neposlední řadě i reklamu. I tak jich zůstalo dost, jak ukazuje třeba pěkný (nákupní) přehled **ActiveStats** (www.statcon.de/statconshop/product_info.htm?language=en&cPath=9_23&products_id=437).

Z hlediska tvorby a prodeje paketů stále dominují USA (i díky státní podpoře), v Evropě se příliš mnoho neděje. Komerčně se z evropských paketů udržel pouze britský **UNISTAT**. Zajímavé jsou německé výzkumné projekty, zejména paket **XploRe** (www.xplore-stat.de/) Humboldtovy university. Přes velký finanční doping se mezi nejúspěšnější nedostaly.

Klíčové produkty lze nalézt u dvou firem:

SAS (www.sas.com/)

resp. **české zastoupení** (www.sas.com/offices/europe/czech) nabízí jeden z nejstarších a nejrozšířenějších paketů **SAS/STAT** (www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/index.htm) nyní ve verzi 9.1 s množstvím nových tajemných metod (viz. **podpora**

(www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/index.htm)). Metod je ovšem tolik, že od verze 8 se pro výuku doporučuje hodně redukovaný **Analyst Application** (vypadá skoro stejně jako SPSS).
Novinkou je také **Akademický program,** který poskytuje software prakticky zdarma pro studenty i učitele (mohou si ho údajné vzít i domů). Ovšem pouze pro silné a bohaté (ze stovky škol je v programu osm).

SPSS (www.spss.com/)
Jeden z nejstarších a nejrozšířenějších paketů. Aktuálně ve verzi 13, na září je ohlášena verze 14, která u nás zatím ke koupi není, ale lze stáhnout demo verzi (nejde mi spustit). Podle velikosti komprimovaného dema přibylo 70% kódu, bude toho asi hodně.
Nový je i **SPSS server 14** a dataminingová **Clementine 9**, přibyl i SPSS Classification Trees 14 ale jenom pro severní ameriku.

Megafirmy SAS a SPSS skoupily řadu dalších produktů, které se prodávají odděleně od hlavního směru. Jsou to zejména kdysi velmi populární pakety

JMP (www.jmp.com/)
Paket původně pro Apple koupený firmou SAS. Od listopadu verze 6, **Novinek** (www.jmp.com/product/jmp6_newfeatures.shtml) je spousta, nic podstatného. Má emauilové novinky i RSS. Také rostoucí sbírku skriptů na nejrůznější témata.

SYSTAT (www.systat.com/)
Oblíbený paket, nyní ve verzi 11. Zlepšena je tradičně kvalitní grafika, novinkou je zejména Monte Carlo/Bootstrap a nové druhy robustní regrese. Stejnojmenná firma nabízí též řadu dalších produktů kdysi koupených firmou SPSS jako **Table Curve**, **Peak Fit** a dalších. (SigmaPlot, SigmaStat, SigmaScan, TableCurve2D, TableCurve3D, PeakFit, AutoSignal) a každý z nich má novou verzi ...

BMDP (www.statsol.ie/bmdp/bmdp.htm)
Oblíbeného paketu se ujala maličká irská firma, která uvedla **BMDP New System Professional 2.0** pro XP. Mají i několik dalších programů pro lékaře.

Další zcela samostatné produkty jsou

STATISTICA (www.statsoftinc.com/)
Je podle vlastního vyjádření hodnocena ve většině srovnání statistických paketů nejlépe.
Současná verze je **STATISTICA 7** a mají už i dataminigový program.
České zastoupení StatSoftu (www.statsoft.cz/) je aktivní a začínají vznikat české popisy, už je i česká STATISTICA 6. Firma pořádá jak kurzy ovládání produktů, tak i obecnější statistické **Kurzy**
(file:///F:/0a_stranky/www.stahroun.me.cz/eseje/statsoft2005/www.statsoft.cz/kurzy)

UNISTAT (www.unistat.com/)
je integrován do prostředí Microsoft Office, což znamená prostředí EXCELu s výstupy do Wordu i HTML.
U nás **UNISTAT.CZ** (www.unistat.cz/), je dokonce **lokalizován do češtiny**.
Současná verze je 5.6, lze si stáhnout hodně očesaný **demo UnistatLight** (u mne chce heslo V55Unistat01November2002) a český manuál. Pracuje pouze s vlastními daty.
Nejnovější verze **light** je hodně očesaná a pro 50 údajů za 8750+DPH je nesmyslně drahá.

STATGRAPHICS Plus (www.statgraphics.com/)

patřil ve své DOSovské podobě k nejoblíbenějším paketům pro běžnou výuku, Současná verze je **Centurion XV**, vypadá parádně.
Bohužel jsem nenalezl české zastoupení a pěkný paket pro výuku zřejmě u nás vyhyne.

GAUSS (www.aptech.com/)
je zejména vhodný pro ekonometrické výpočty. Současná verze má číslo 7.

MINITAB (www.minitab.com/)
Stále ve verzi 14.2., novinky pouze ze života. Lze stáhnout **demo verzi**. Součástí je soubor tutoriálů a pomocniků KeepingTab.
ActivStats for MINITAB je interaktivní miltimediální statistický text, ve kterém jsou animace, dynamické grafy a videoklipy.

STATA (www.stata.com/)
Paket inovoval na verzi 9 a trochu zdražil. Nový maticový jazyk, mnoho statistiky, **podpora 64 bitových procesorů**.

S-PLUS (www.insightful.com/),
U nás podporuje tento paket firma **TriloByte** (www.trilobyte.cz/),
Veliký pokrok. Verze 7 pracuje s obrovskými soubory, implementuje pokročilou statistiku atd. Také nový **Enterprice Server 14**. Původně určen a využíván akademickými profesionály se zjevně snaží vetřít do velkých korporací.

WINKS (www.texasoft.com/)
malý, jednoduchý a levný (od 59 dolarů) paket, naprosto postačující pro výuku. Nyní verze 4.8. On-line manual, tutoriály metod atd. **Neinovoval!**

NCSS (www.ncss.com/)
Rozsáhlý paket je ve verzi NCSS 2004, v září 2005 vyšel upgrade (zdarma).

GENSTAT (www.vsn-intl.com/)
britský paket s dobrou podporou, inovoval na verzi 8 a přidal SP1. Je nyní **zdarma pro rozvojové země**.

3. Tabulkové procesory

Dominuje **EXCEL**. Je lokalizován, je všeobecně dostupný a víceúčelový, obsahuje v podstatě většinu základní statistiky včetně práce s daty, grafiky a návaznosti na publikační software.

Na bázi EXCELu lze vybudovat celý statistický paket nebo aplikace, mnohé pakety obsahují rozhraní přímo z EXCELu (zejména ty velké).

Existuje k němu i řada statistických nadstaveb (třeba UNISTAT, který je dokonce lokalizován do češtiny) a existují i nadstavby zdarma (pěkný je **MERLIN** (www.heckgrammar.kirklees.sch.uk/content/departments/science/biology/merlin.htm)).

I u nás se objevují hotové statistické skripty pro EXCEL, letos zejména **Klára Mrázová** (www.pcsvet.cz/art/author.php?id=429&page=2).

4. Matematické systémy

Matematické systémy začínají dnes být vážným konkurentem statistických paketů v oblasti výuky. Přesto, že jsou mimo USA velmi drahé, zaměřují se na střední školy v naději na úspěch později.

Nejznámější jsou zejména systémy:

MATHEMATICA (www.wolfram.com/)

(u nás **Elkan** (www.elkan.cz/)) Novinkou je verze 5.2, která zejména reaguje na vývoj v oblasti procesorů (64 bitové a vícejádrové procesory).
Pokračuje i integrace statistiky, ale přidali nyní jenom **stem and leave diagram**.
Pro méně náročné a movité přibyl nový **CalcCenter3** (www.wolfram.com/products/calccenter/why/mathematica.html) se zjednodušeným ovládáním a podstatným zrychlením.
Na firemní stránce přibylo mnoho aplikací zdarma od uživatelů, např. **eFunda** (www.efunda.com/webM/home/webM_home.cfm) ale jsou jich tisíce z mnoha oborů.

MATLAB (www.mathworks.com/)

Velmi rozšířený matematický paket, jeho souputník **SIMULINK** prakticky dominuje v simulacích. U nás jej distribuuje firma **Humusoft** (www.humusoft.cz/indexcz.htm).
V současné době je ve 14. vydání, která přineslo kromě velkého zrychlení i revoluční možnost kompilace výpočtu (lze je použít nezávisle na paketu), generátor grafů, zlepšení výstupů atd. Inovací je velmi mnoho, realizují se zejména prostřednictvím Service Packů (zatím 3, poslední 1.11)
SP 2 přinesl **paralelní výpočty** a **Bioinformatics Toolbox** pro grafické a statistické zpracování měření.
SP 3 kromě jiného také pár statistických grafů.
Průběžně inovují i spousty toolboxů, pro statistiky je zajímavý inovovaný **Statistics Toolbox 5.1** (www.mathworks.com/products/statistics/), který je zdarma.
Existuje také celé řada klonů MATLABU zdarma, viz **Mathlab Clones** (www.dspguru.com/sw/opendsp/mathclo2.htm)

Mnoho přidávají uživatelé, často zdarma. Třeba **Datatool** (www.datatool.com/prod01.htm) vizualizační software pro Matlab, zdarma pro akademické potřeby.
Na novinky má **RSS** (www.mathworks.com/company/rss/index.html?fp).

MATHCAD 13 (www.mathcad.com/products/mathcad13/)

(též **Mathsoft.cz, české zastoupení** (www.mathsoft.cz/)) je každoroční inovace příjemného inženýrského matematického paketu který má též miliony uživatelů a rozsáhlé statistické možnosti. Přirozený zápis matematických struktur je vhodný pro výuku a částečně i díky výhodným cenám proniká i na střední školy.

5. Internetové interaktivní statistické učebnice

Velké zklamání, nic nového se zřejmě neděje.

6. Open source, freeware, shareware

Freeware

je zdarma, jedná se obvykle o méně rozsáhlé projekty s omezenou životností. John Pezzullo je autorem rozsáhlého přehledu statistického softwaru zdarma. **Javastat** (www.statistics.com/content/javastat.html).

Open source

software je nejenom zdarma, ale je zveřejněn i programový kód a možno jej měnit. Výsledkem je obvykle dynamický vývoj a některé projekty jsou velmi úspěšné (třeba Linux, Mozilla, PHP).
Z nejzajímavějších produktů lze jmenovat
The R Project for Statistical Computing (www.r-project.org/)
Jazyk **R** je open source klon **S+** s velmi širokou škálou navazujících produktů a dynamickým vývojem. Současná verze je 2.2.0 ze září 2005 (novější než na CD Společnosti).
Nové verze by měly vycházet vždy 1.4. a 1.9., lze se zúčastnit vývoje na beta verzích a objednat časopis novinek.

SciLab (scilabsoft.inria.fr/)
trochu okleštěná freewarová verze Matlabu je nyní také open source projekt! Nyni verze 3.1.1. Spolupracuje i s freewarovou verzí matematického paketu **MuPad** (research.mupad.de/).

Shareware
je na zkoušku a levně, finančně v možnostech jednotlivce.
Úroveň je různá, ale nápady velmi často výborné a s trochou trpělivosti lze najít všechno. Přesto se ve statistice se příliš nepoužívají.
Shareware a specializované pakety obvykle nemají adekvátní podporu a mnoho lidí jim prostě nerozumí. Je i malá důvěra ve spolehlivost výsledků a neochota se učit nestandardní ovládání.

7. On-line výpočty zdarma

Jejich význam roste vzhledem k růstu dostupnosti širokopásmového internetu. Je jich mnoho, často dostupných pouze studentům a učitelům jediné školy či instituce.

Několik veřejných:

Rweb (www.math.montana.edu/Rweb/)
Internetové rozhraní k jazyku R. Existuje několik dalších serverů, na kterých lze přímo z internetu počítat v jazyce R (Rweb nepokrývá vše, pouze základní moduly).
Chtělo by to něco podobného u nás.
Free Statistics and Forecasting Software (www.wessa.net/)
několik komplexních výpočtů

Internetové rozhraní je jedním z nejperspektivnějších směrů nejenom pro pakety, ale veškerý statistický software.

8. Specializované programy

Sem zařazujeme zpravidla **dílčí statistické systémy**, které pokrývají pouze určité statistické metody. Je jich obrovská spousta pro nejrůznější úlohy a často to jsou astronomicky drahé speciální programy pro bohaté firmy a úřady. Mnoho statistických firem si vytváří vlastní software.

Vystopovat lze zejména některé okruhy šířeji použitelných metod, které mají vlastní statistické programy a pakety. Dynamický vývoj je zejména ve dvou oblastech

Dataminig
: Datamining je tak trochu paběrkování na datových smetištích, kde vyslovovat klasické předpoklady o náhodných veličinách se zdá být neadekvátní, takže klasické statistické usuzování jde trochu stranou. Metody jsou spíše heuristické, fundamentalističtí statistikové trochu ohrnují nos, takže se převážně zařazují do informatiky.
Metody bývají jako samostatné programy u všech významnějších firem. Asi nejpokročilejší je asi nová SPSS Clemetine 9.
Novinkou roku je **text mining**.

Statistické grafy
: Význam grafiky obecně stoupá, možnosti počítačů a požadavky na estetické ztvárnění se zvyšují. Klasické statistické pakety se sice zlepšují, ale v konfrontaci se současnými možnostmi interaktivní dynamické grafiky jsou hodně dlužny.
Toho využily spousty firem specializovaných na **statistickou grafiku**. Novinek je příliš mnoho.

Z ostatních novinek mne zaujaly zejména:

Analýza přístupu na stránky
: je nová oblast aplikací statistiky, kde je mnoho produktů a dokonce existují výukové kurzy (Navrcholu).
Novinkou jsou zejména **Google Analytics** (www.google.com/analytics/), které jsou v atraktivní grafice a zdarma. Mohou sloužit jako krásný příklad při výuce popisné statistiky!

SSI (www.ssicentral.com/index.html)
: má novou stránku a nejslavnější program LISREL (strukturální modely) je ve verzi 8.72. Starší verze pro výuku zdarma.

Xtremes Webpage (www.xtremes.math.uni-siegen.de/xtremes/)
: Stránka ke knize Extremální rozdělení a specializovanému programu (starší verze pro výuku zdarma). Nová kniha (3.vydání, vyjde tuto zimu), program 4.0, StatPascal.

Resampling Stats (www.resample.com/)
: Samotný program už ve verzi 5.0, ad-ony k EXCELu a MATLABu. Metody přestavby výběru a simulace jsou na vzestupu.

Specializované programy sice odebírají paketům významnou část trhu, ale zejména v oblasti výuky je plně je nahradit nemohou.

9. Závěr?

Skoro stejný jako vloni. Nicméně co nás čeká a co by se dalo dělat. Hlavní problém je výuka, týká se zřejmě největšího počtu lidí.

Ceny počítačů dramaticky v čase klesají (byl představen **100$ laptop** (www.physorg.com/news8255.html)) a dostupnost internetu rychle roste a dostává se i na mobily. Bude potřeba udělat **pakety pro mobily** (mají obrazovku, paměť, Javu).

Cena statistických paketů spíše roste.

Přitom už 46 let je základní rozsah používaných a vyučovaných metod definován a naprogramován. Stále se něco děje spíše v oblasti ovládání a prezentace. Přibývají sice stále nové a nové metody, ale jsou vysoce specializované a ani odborníci to už nestačí sledovat. Pokrok existuje, měřitelný spíše velikostí (třeba 30 denní komprimovaný SPSS Eval má v jednotlivých verzích 11.5(2002): 57MB, 12:1004MB(2003), 13(2004):109MB, 14(2005):160MB) je zcela obdobný ostatnímu software a stěží opravňuje růst cen, když užitná hodnota zůstává pro většinu uživatelů stejná.

Kvůli ceně se ani při výuce nepoužívají jednotným způsobem statistické pakety a často se používá pouze EXCEL.

EXCEL stačí pro běžné elementární úlohy, má skvělou práci s daty, je lokalizován a vše základní lze spočítat. Stačí pro střední ekonomické kádry, které jej stejně potřebují, ale vytrácí se statistické myšlení.

Obecně nasazení EXCELu nemusí být špatně, protože většina paketů dnes stejně spolupracuje s EXCELem a lze přidat i statistickou nadstavbu (třeba UNISTAT, ADSTAT ale jsou i zdarma). Bastlení výpočtů pouze v EXCELu je hrůza.

Snad by pomohly skripty. Vyzývám k **centrální databázi statistických skriptů** pro EXCEL pro běžnou statistickou výuku (na internetu už takovéto skripty jednotlivě jsou).

Přibývá výuka statistiky bez výpočtů i paketů.

Výsledkem je **ekostatistika**, výuka statistiky prováděná metodami předminulého století, nejednotná, nekontaminovaná vzorci a statistickým softwarem. Zejména u nás nabývá hrozivého rozsahu na minivysokých školách.

Rozkvět ekostatistiky je varující. I pro statistické softwarové firmy, je i v jejich zájmu aby byli lidé statisticky gramotní. Nesmí vzniknout intelektuální hranice mezi bohatými a chudými (zvlášť když jsou v USA pakety mnohem levnější).

Podle mého názoru je více možností

Minipaket
: obsahující pouze skutečně potřebné metody pro základní výuku s omezením rozsahu dat na výuku a pěknou grafikou.
Snad nejblíže tomu je **STADIA 6.2** (www.exponenta.ru/soft/Others/stadia/stadia.asp), rozšířený ruský výukový paket zdarma (v českých Windows vypíná počítač už při instalaci).
Možnost se nabízí, třeba diplomka u informatiků jako základ **Open source projektu**.
V Čechách i na Slovensku je dost programátorů, kteří s tím mají zkušenosti.
Problém je v tomto případě u statistiků, pokroky v informatice je nezajímají a nejsou schopni spolupracovat, a to i když jsou na stejné fakultě.

On-line rozhraní
: Firmy by místo obrovských demo souborů na 30 dní, které omezují pouze slušné lidi (lump to crackne), poskytnout přístup k programu pro rozumně omezená data případně metody.
Jinak je to hotovo, dělají to **statistické servery** (ale to už bývá pro zvané).

Až bude statistický software rozumně dostupný jednotlivcům, tak si masy na statistický software zvyknou a nezbude jim nic jiného než si koupit příslušný ad-on ke známému produktu pro komerční úlohy.

Prostě se trochu poučit třeba od Microsoftu, který nabízí masově produkty v rozumných cenách (viz třeba nyní redukované Windows XP pro Čínu, Indii a Rusko) a poslední době uvažuje i o **Windows zdarma** (www.novinky.cz/internet/69943-microsoft-mozna-poskytne-windows-zadarmo.html) a bohatne ze služeb

Trh v klasických zemích už moc neporoste, tisíce prodaných paketů nikoho neuživí. Paketů lze prodat tisíckrát více za setinu ceny a vydělat 10 krát více.

Internetovská, průběžně inovovaná verze je na **zde**
(www.stahroun.me.cz/eseje/statsoft2005/index.htm).

Autor:
Doc.Ing.Jiří Žváček, CSc.
V úvalu 84, nem.Motol, LDN,7.stanice
15000 Praha 5
jzvacek@seznam.cz

PREHLIADKA PRÁC MLADÝCH ŠTATISTIKOV A DEMOGRAFOV

Modelovanie rizikových faktorov vedúcich k indikácii antibiotickej liečby u akútnej angíny pomocou logistickej regresie

Katarína Baňasová

Súhrn
Cieľom práce bolo vyvinúť predikčné pravidlá pre indikáciu antibiotík pri liečbe akútnej tonzilofyringitídy. Dáta pochádzajú z preskripčnej štúdie Infekcie dýchacích ciest, ktorá prebehla v Novembri 2003 v ambulantnej praxi 66 lekárov na Slovensku. Dátový súbor obsahoval hlavné charakteristiky pacienta, ich rizikové faktory a iné záznamy o liečbe. Uvažovali sme o 15 premenných, ktoré môžu ovplyvniť preskripciu antibiotík a ich vzájomné interakcie. Použili sme model logistickej regresie, ktorej odhady parametrov sme urobili pomocou programu SAS Enterprise Guide. Výsledný model obsahoval šesť premenných.
Kľúčové slová
Logistická regresia, antibiotická liečba, akútna tonzilofaryngitída.

Summary
The aim of this work was to develop a prediction rule for antibiotic indication at the treatment of acute tonsillofaryngitis. Data were from prescription study of Acute Respiratory Tract Infections, which passed in November 2003 in 66 primary practices in Slovakia. The data file consists of main characteristic of patients, their risk factors and other records of the therapy. We considered the 15 risk factors and their interactions. We used a logistic regression in the SAS Enterprise Guide. The model was built on the six variables, which mainly affect an antibiotic indication at this infection.
Key words
Logistic regression, Antibiotic treatment, Acute tonsillofaryngitis.

1. Úvod a metódy

Logistická regresia je metóda, ktorou získame odhady podmienenej pravdepodobnosti jednej z hodnôt kategoriálnej závislej premennej od iných nezávislých premenných. Ako nezávislé premenné (prediktory) sa môžu použiť tak kategoriálne ako aj spojité premenné, pričom tieto nezávislé premenné nemusia spĺňať predpoklad normálneho rozdelenia.

Na odhad modelu logistickej regresie sme použili dáta z medicínskeho výskumu. Sú to dáta popisujúce rozhodovanie lekára pri liečbe akútnej tonzilofryngitídy (angíny), infekčného ochorenia horných ciest dýchacích. Súbor obsahuje 4664 záznamov liečby. Modelovali sme binárnu premennú podanie antibiotika s názvom *Bez_ATB*, jej hodnotu „n“(n=podanie, a=nepodanie). Uvažovali sme s 15 nezávislými premennými jednotlivo a tiež v ich vzájomných interakciách. Na výber významných premenných do modelu sme použili metódu Stepwise selection. Táto metóda bola vykonaná na hladine významnosti 0,05 pre zaraďované aj vyraďované premenné. Do výsledného modelu sa dostalo šesť štatisticky významných premenných, u ktorých bola p-hodnota menšia ako hladina významnosti 0,05 (Tabuľka 1) a do modelu nebola zaradená žiadna interakčná premenná.

Tabuľka 1. Výstup procedúry Stepwise Selection

Summary of Stepwise Selection					
Step	**Effect Entered**	**DF**	**Score ChSq**	**Pr > ChSq**	**Popis hodnôt premenných**
1	**Etiológia**	2	1842,75	<0,0001	b=bakteria, n=nejasná, v=vírusová
2	**Pomoc_vyš** (PV)	1	65,5916	<0,0001	0=nie, 1=áno
3	**Vek_interval**	4	17,2354	0,0017	1=0-5; 2=5-10;3=10-15;4=15-20;5=20-25 rokov
4	**Mikrob_vyš**	1	9,2868	0,0023	0=nie, 1=áno
5	**Kolektiv_zariade**	1	7,027	0,008	0=nie, 1=áno
6	**Po_hospitalizácii**	1	4,5341	0,0332	0=nie, 1=áno

2. Výsledky

Bodový odhad výsledného logistického modelu má nasledovnú rovnicu:

$$Logit(p_i) = -3.03+4.71\ A_{Etiologia\_b}+3.10\ A_{Etiologia\_n}+ 0.12\ A_{vek\_sk}+0.81\ A_{PV}+ 0.34\ A_{MV} -0.3216\ A_{Kolektív}+ 1.09\ A_{Po\_hospit}$$

Model ako celok je štatisticky významný a jeho vierohodnostný pomer a ostatné štatistiky ako je Akaikeovo informačné kritérium (AIC) a Schwarz-Bayesovo informačné kritérium (SBC) sú uvedené v tabuľke č. 2. Sú to penalizačné miery kvality odhadu, ktoré merajú presnosť modelu, ale zároveň ju penalizujú počtom parametrov, ktoré boli potrebné na dosiahnutie tejto presnosti. Možno ich použiť na porovnanie konkurenčných modelov. Vidíme, že odhadnutý logistický model s 6 prediktormi má uvedené štatistiky omnoho nižšie ako by mal model len s lokujúcou konštantou.

Tabuľka 2. Miery kvality vyrovnania

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and covariates**
AIC	5353,1	3497,154
SC	5359,531	3548,606
-2 Log L	5351,1	3481,154

Testovacia štatistika pri logistickej regresii vychádza z dvoch mier: upravenej vierohodnostnej štatstiky -2 LOG L = -2 Σ log (pi) a zo štatistiky Score. Pre obidve miery sa počíta Waldov chí-kvadrát testovacia štatistika a jej zodpovedajúca p-hodnota. Pre prijatie alebo zamietnutie hypotézy platia nasledujúce pravidlá: ak $p > \alpha$ prijímame hypotézu H_0, teda parameter nie je významný; ak $p < \alpha$ nezamietame hypotézu H_1, a teda parameter je významný. V našom prípade p-hodnoty sú $p < \alpha$, teda vybrané parametre sú významné.

Tabuľka 3. Waldova chí-kvadrát štatistika

Type 3 Analysis of Effects			
Effect	**DF**	**Wald Chi-Sq**	**Pr > ChiSq**
Etiológia	2	1098,793	<0,0001
Kolektív_zariade	1	6,6088	0,0101
Mikrob_vyš	1	8,937	0,0028
Po_hospitalizácii	1	4,4476	0,035
Pomoc_vyš	1	45,4247	<0,0001
Vek_interval	4	21,3585	0,0003

Pomer šancí (Odds ratio) je štatistika, ktorá určuje s ohľadom na náhodu o čo väčšiu resp. menšiu máme šancu, že sa konkrétna obmena závislej premennej vyskytne v jednej zo skupín v porovnaní so šancou jej výskytu v inej skupine. Je nástrojom na interpretáciu modelu. Pri bakteriálnej etiológii je šanca, že bude antibiotikum podané 115 ku 1 vyššia ako pri vírusovej etiológii (Tab. 4.). Taktiež to platí pre nejasnú etiológiu, pri ktorej lekár častejšie antibiotikum indikuje. Parameter, ktorý popisuje návštevu kolektívneho zariadenie (napr. školského) vplýva na podanie antibiotika negatívne, tzn. ak dieťa nenavštevuje kolektívne zariadenie, je šanca väčšia, že mu bude antibiotikum podané.

Tabuľka 4. Pomer šancí, bodové odhady pravdepodobností a 95% Waldove intervaly spoľahlivosti

Odds Ratio Estimates			
Effect	**Point Estimate**	**95% Wald Confidence Limits**	
Etiologia b vs v	115,719	87,358	153,287
Etiologia n vs v	22,931	17,884	29,403
Kolektiv 0 vs 1	1,424	1,088	1,863
MV 0 vs 1	1,386	1,119	1,716
Po_hospital 0 vs 1	0,337	0,122	0,926
PV 0 vs 1	2,236	1,77	2,826
V_sk5 1 vs 5	1,024	0,25	4,197
V_sk5 2 vs 5	1,093	0,268	4,465
V_sk5 3 vs 5	1,773	0,433	7,254
V_sk5 4 vs 5	1,266	0,309	5,197

Vykonanie pomocného alebo mikrobiologického vyšetrenia pred zahájením liečby značne prispieva k nižšej indikácii antibiotika pri tomto ochorení. Z rizikových faktorov k indikácii antibiotika ďalej významne prispieva predošlá hospitalizácia pacienta a vek pacienta. Premennú vek sme rozdelili na vekové skupiny v 5-ročných intervaloch. V porovnaní k dospelému pacientovi, pravdepodobnoť, že detský pacient dostane antibiotikum, je vyššia. Šanca indikácie antibiotika sa porovnávala medzi každým vekovým intervalom k intervalu „5“, t.j. k 20-25 ročných pacientom. Pomer šancí je najvyšší pri vekovej kategórii 10-15 ročných pacientov, ktorí sú pri tejto diagnóze najčastejšie liečení antibiotikami.
Miery asociácie určujú asociáciu medzi predikovanými pravdepodobnosťami pre obmenu závislej premennej a skutočnými hodnotami závislej premennej. Čím vyššia je asociácia medzi týmito hodnotami, tým je model kvalitnejší pri predikcii závislej premennej. Vychádzajú z párovania pozorovaní. V tabuľke č. 5 sú uvedené miery asociácií párov predpovedaných a nameraných hodnôt. Náš model predpovedal zhodné páry na 85,2%, nezhodné páry na 12,7% a zmiešané páry na 2%. Odvodené miery asociácie, ako Somersovo D, Gamma, Tau-a a „c“, nadobúdajú hodnotu od 0 po 1. Model je tým kvalitnejší, čím sú hodnoty odvodených mier bližšie k 1. Náš model dosahuje v týchto ukazovateľoch hodnoty od 0,286 po 0,863.

Tabuľka 5. Odvodené miery asociácie

Association of Predicted Probabilities and Observed			
Responses			
Percent Concordant	85,2	**Somers' D**	0,725
Percent Discordant	12,7	**Gamma**	0,74
Percent Tied	2	**Tau-a**	0,286
Pairs	4148538	**c**	0,863

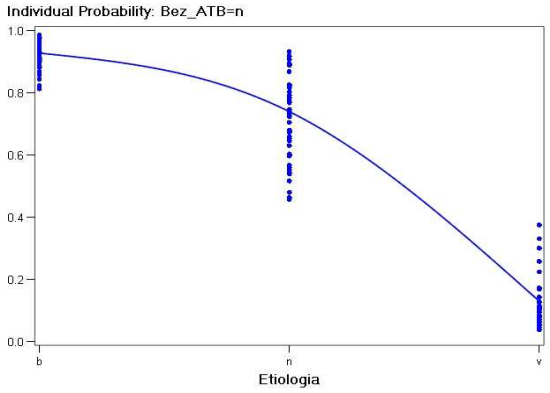
Pri priebehu procedúry logistickej regresie sa porovnávajú namerané a predpokladané hodnoty predikovanej premennej. V novovytvorenom dátovom súbore sú tieto hodnoty uložené v stĺpci FROM a INTO. Tabuľkovou analýzou sme vytvorili tabuľku 6, v ktorej môžeme porovnať správne a mylne predikované páry. Model správne predpovedal podanie antibiotickej liečby u 3228 prípadov (96,32% z pôvodne nameraných hodnôt podania antibiotika). Antibiotickú liečbu by na základe parametrov neindikoval u 681 prípadov (55,01% z prípadov bez antibiotickej liečby). Teda celkovo správne predpovedal 3909 prípadov (85,18%).

Tabuľka 6. Tabuľková analýza pôvodných a predikovaných hodnôt.

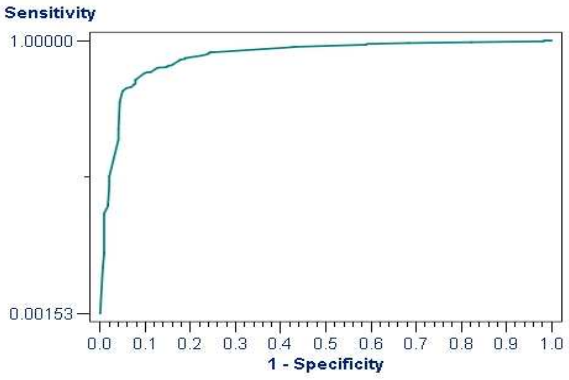
Table of _INTO_ by _FROM			
INTO	**FROM**		**Total**
	a	**n**	
a	**681**	123	804
n	557	**3228**	3785
Total	1238	3351	4589

Nelineárnu závislosť medzi pravdepodobnosťou podania antibiotika a hodnotami nezávislej premennej etiológia infekcie zobrazuje logistická krivka (Obr. 1.). Stanovenie pôvodcu infekcie je najdôležitejším faktorom, ktorý ovplyvňuje indikáciu antibiotika. Lekár pri predpokladanej vírusovej infekcii (v) nemá dôvod indikovať antibiotikum. Liečba by bola neadekvátna a neúčinná, preto pri vírusovej etiológii sú hodnoty predikovanej pravdepodobnosti nízke (0,4-0,0). Pri nejasnom pôvode infekcie (n), tzn. lekár nevie stanoviť, či ide o bakteriálneho pôvodcu alebo vírusového, prevláda z rôznych dôvodov podanie antibiotika. Hodnoty pravdepodobnosti varírujú od 0,45 do 0,95. Pri bakteriálnej etiológii ide takmer o 100% podanie antibiotík (pravdepodobnosť od 0,81-0,99). Popisovaný graf závislosti má v optimálnom prípade tvar S-krivky. Vzhľadom k tomu, že model obsahuje aj iné prediktory hodnoty pravdepodobnosti podania antibiotika pre jednotlivé hodnoty etiológie sú v zobrazených intervaloch. Interpoláciou stredových bodov získavame odhad S-krivky.

ROC krivka (Receiver Operating Characteristic Curve) zobrazuje senzitivitu a špecificitu modelu. Platí tu, že čím je krivka viac vzdialená od diagonály, tým je model špecifickejší a citlivejší. Náš model je dostatočne citlivý na zmenu dát a špecifický k vybraným parametrom.



Obrázok 1.: Odhad pravdepodobnosti modelovanej premennej v závislosti od etiológie.



Obrázok 2.: ROC krivka.

3. Záver

Logistickou regresiou sme vytvorili model, ktorý na základe rôznych rizikových faktorov liečby akútnej tonzilofaryngitídy predikuje podanie antibiotika na 86,3%. Zo všetkých uvažovaných parametrov k indikácii antibiotika vedie najmä predpokladaná bakteriálna etiológia ($p<0,0001$). Nevykonanie pomocného ($p<0,0001$) a mikrobiologické vyšetrenia ($p=0,0028$) taktiež vo veľkej miere ovplyvňuje indikáciu antibiotík. Ich použitím sa exaktne stanoví typ infekcie, ktoré umožní predísť častému a neadekvátnemu podaniu antibiotika pri vírusovej infekcii. Ďalším parametrom je vek pacienta. Najvyššia pravdepodobnosť antibiotickej liečby bola u pacientov vo veku 10-15 rokov. Z rizikových faktorov pacienta bol najvýznamnejší parameter návšteva kolektívneho zariadenia a stav po hospitalizácii.

Zdroj dát

Preskripčná štúdia Infekcie dýchacích ciest November 2003 (Hupková, H., Foltán, V.)

Literatúra

[1] Bont, J., Hal, E., Hoes, A.W., Bonten, M. J. M. et al.: Development of a prediction rule for hospitalisation or death in elderly primary care patients with a lower respiratory tract infection. In: Clinical Microbiology and Infection. - ISSN 1198-743X. - Vol. 11, Suppl. 2, 2005, p. 429.

[2] SAS OnLine Documentation: http://support.sas.com/91doc/docMainpage.jsp

[3] Sharma, S.: Applied Multivariate Techniques. New York, John Wiley and Sons, Inc., 1996.

[4] Stankovičová, Iveta: Logická regresia a hĺbková analýza dát: (Data Mining). In: Štatistické metódy v praxi. Bratislava: SŠDS 2002. ISBN 80-88946-19-0

[5] Vojtková, Mária: Odhad parametrov logistickej trendovej funkcie pomocou regresie. Ekomstat ’90. 3. škola štatistiky. Zborník príspevkov. Trenčianske Teplice, s. 22-24, Bratislava: SŠDS, 1990.

Adresa autorky:
Katarína Baňasová, Fakulta managementu UK, 3 ročník
Katedra riadenia a organizácie farmácie, FaF UK, Kalinčiakova 8, 832 32 Bratislava
Email*: Katarina.Banasova@St.Fm.Uniba.sk*

Kvantifikácia determinantov výšky bankových vkladov

Ján Beracko

Abstrakt

It is supposed that the whole amount of resident's bank deposits in domestic currency depends on resident's income, gross domestic product, nominal interest rate of deposits, inflation rate, exchange rate of Slovak crown and euro, alternate pay-off (index SAX) and amount of resident's bank deposits of the preceding period. The main accent was given on statement of significance of set variables and estimation of their determinants using the Ordinary Least Squares Estimators. The output of given analysis has shown the significance of four variables, namely resident's income, gross domestic product, nominal interest rate and amount of resident's bank deposits of the preceding period.

Úvod

Peniaze v dnešnej dobe majú, z pohľadu ich funkcií dôležité postavenie, pričom nie je dôležité v akej podobe sa v danej ekonomike a čase vyskytujú. Vo všeobecnosti existujú v trhovej ekonomike dvaja emitenti peňazí: centrálna banka, ktorá emituje hotové (hotovostné) peniaze a komerčné banky, ktoré emitujú depozitné peniaze, vznikajúce ako poskytovanie úverov prostredníctvom komerčných bánk. Centrálna banka tento proces ovplyvňuje pomocou monetárnych nástrojov – miera povinných minimálnych rezerv, operácie na voľnom trhu, základná úroková miera. Bankový systém založený na ukladaní časti vkladov v podobe povinných minimálnych rezerv, vytvára priestor pre tvorbu depozitných peňazí. Schopnosť niekoľkonásobného zvýšenia bankové vklady sa vyjadruje prostredníctvom multiplikátora depozít (vkladov).
Cieľom práce je overenie predpokladu modelovania bankových vkladov aj bez menovej bázy. Podľa ekonomickej teórie by mala veľkosť bankových depozít (vkladov) na účtoch domácností byť úzko prepojená s rastom HDP, príjmami obyvateľstva, infláciou a nominálnou úrokovou mierou bankových vkladov, pri racionálne sa správajúcich klientoch aj s kurzom slovenskej koruny a zahraničnej meny (euro) a s alternatívnym výnosom napr. z akcií (hodnoty indexu SAX).
Po kvantifikácii determinantov a ich vplyvov na vysvetľovanú premennú a po overovaní vhodnosti modelu sa pokúsime zistiť, či je výsledný model vhodný aj na konštrukciu predpovedí.

Dáta a metodika

Analýza sa uskutočnila v prostredí slovenskej ekonomiky za obdobie od 4. kvartálu roku 1995 do 2. kvartálu roku 2005, pričom metódou použitou pri určovaní determinantov výšky bankových vkladov bol odhad ekonometrického modelu metódou najmenších štvorcov[1] v prostredí ekonometrického softvéru Eviews.
V základnom tvare sa dá predmet analýzy zapísať ako funkčne závislá premenná od vysvetľujúcich premenných nasledovne, BV≈ *f*(P, NUM, INF, HDP, EUR, SAX), kde BV sú

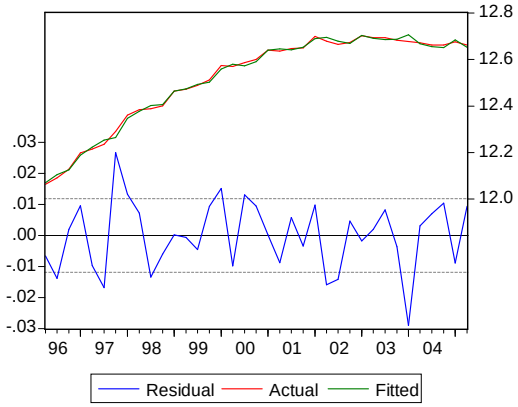
[1] OLS – Ordinary Least Squares

termínované a netermínované vklady obyvateľstva v bankách v domácej mene (SKK), P je príjem domácností, NUM je nominálna úroková miera bankových vkladov v domácej mene (SKK), INF je miera inflácie, HDP je hrubý domáci produkt, EUR je výmenný kurz eura a slovenskej koruny a SAX je alternatívny výnos v podobe akciového indexu SAX.
Prvotný model vychádzal z teoretických predpokladov, ktoré boli spomenuté v úvode. Bol doplnený o predpoklad závislosti BV na výške BV z minulého obdobia, a o legovanie P a NUM, kvôli rozhodovaniu sa domácností pod vplyvom minulých hodnôt týchto premenných.
Z počiatočného modelu boli vyradené nevýznamné premenné a po grafickej analýze vzájomnej závislosti premenných sme niektoré premenné zlogaritmovali. Po zlogaritmovaní P sa zvýšila významnosť premennej NUM v podobe, v akej sa vyskytuje v modeli, t.j. pre BV_t to je rozdiel $NUM_{t-1} – NUM_{t-2}$ (diferencia - $D(NUM_{t-1})$. Týmito úpravami sa zvýšila presnosť modelu vo významnosti jednotlivých determinantov a aj modelu ako celku, čo demonštrujú t-štatistiky a pravdepodobnosť nevýznamnosti determinantov a vylepšené ukazovatele R^2 (0,996743) a $R^2_{adj.}$ (0,996336), zníženie informačných kritérií (AIC = -5,902466 a SIC = -5,684774) a aj nasledujúci obrázok, na ktorom je vidieť úbytok rozptylu rezíduí.

Tabuľka 1 Odhady regresných koeficientov

premenná	koeficient	štandardná chyba	t - štatistika	pravdepodobnosť
C	4,354555	0,470625	9,252703	0
LOG(P(-1))	0,253247	0,031958	7,924308	0
LOG(HDP)	-0,473698	0,062412	-7,589905	0
D(NUM(-1))	0,009354	0,003609	2,591687	0,0143
LOG(BV(-1))	0,865094	0,02359	36,67257	0

Z grafu znázorňujúceho komparáciu vývoja skutočných hodnôt a odhadovaných hodnôt je zrejmé, že model pomerne dobre popisuje pôvodnú funkciu (pomerne dobre kopíruje skutočnosť). Najvýraznejšie problémy nastávajú pri popise vývoja na konci roka 1997 a prelome rokov 2003 a 2004.



Obrázok 1

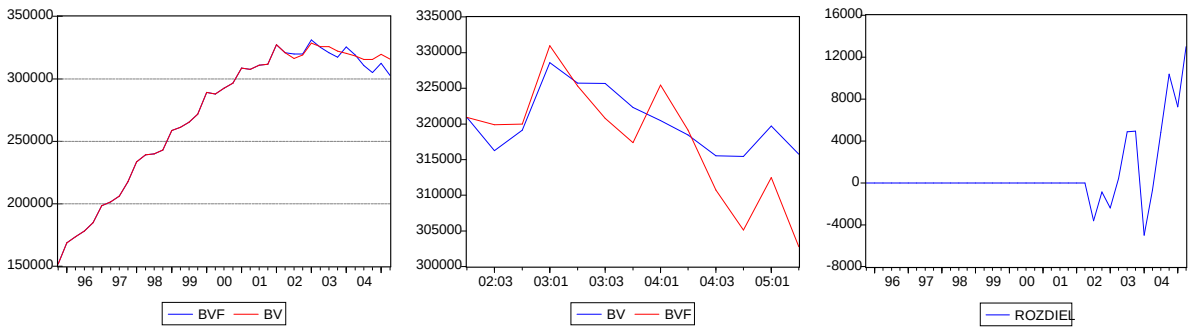
Zistenia

Vynechanie vysvetľujúcich premenných alebo chybná špecifikácia analytického tvaru modelu sa na základe Ramseyho RESET testu nepotvrdila. Testuje možné pridanie vysvetľovanej premennej do funkčného modelu ako regresora v tvare zodpovedajúcom kvadratickému členu, resp. kvadratickému funkčnému tvaru modelu. Výsledok však potvrdzuje správnosť konečného modelu – nie potrebné pridávať dodatočnú premennú do modelu.

Heteroskedasticita ako porušenie požiadavky konečného a konštantného modelu sa na základe Whiteovho testu nepotvrdila, keďže R^2 dosiahlo len malú hodnotu 0,228015. Na základe hodnôt z korelačnej matice a následne z výstupu prvotného modelu, na základe ktorého boli vyradené niektoré premenné, kde mohlo ísť o multikolinearitu (vysoké korelácie medzi premennými), sa zdá, že medzi jednotlivými premennými konečného modelu sa už tento neželaný jav nevyskytuje. Na detekovanie autokorelácie kvôli obsahu oneskorenej závislej premennej medzi vysvetľujúcimi premennými nebol použitý Durbin-Watsonov test autokorelácie rezíduí. Na rozdiel od Durbin-Watsonovej štatistiky LM testy[2] sú aplikovateľné na detekovanie autokorelácie aj v prípade, keď analytický tvar modelu obsahuje oneskorenú závislú premennú. Z LM testov sme sa rozhodli využiť Breusch-Godfreyho test[3], na základe ktorého nemôžme zamietnuť nulovú hypotézu o sériovej nekorelovanosti náhodných chýb daného ekonometrického modelu, v tomto prípade štvrtého stupňa. Testovacia štatistika tohto testu je uvedená v (1).

$$e_t = \mathbf{x}_t^T \boldsymbol{\gamma} + \sum_{s=1}^{4} \alpha_s e_{t-s} + \eta_t \sim \chi^2_{(4)} \qquad (1)$$

Predpovedaciu schopnosť analytického tvaru modelu sme robili pre interval dát od 3. kvartálu roku 2002 do 2. kvartálu roku 2005 a ukázala sa ako dobrá (Obrázok 2).



Obrázok 2
Porovnanie skutočnej výšky bankových vkladov s predikciou

Ako je vidieť z grafov model celkom presne popisuje skutočný vývoj, pričom odchýlka (chyba), ktorá vznikla nie je až taká veľká (Obrázok 2) – predstavuje len niečo okolo 4 mld. Sk, čo pri celkových bankových vkladoch v hodnote pohybujúcej sa okolo 320 mld. Sk (v posledných obdobiach) nepredstavuje veľkú nepresnosť. Theilov koeficient nesúladu taktiež potvrdil dobrú predpovedaciu schopnosť analytického tvaru modelu.

Záver

Daný model nadobudol po nevyhnutných úpravách nasledujúci tvar:

$\log(BV) = 4{,}3546 + 0{,}2532*\log(P(-1)) - 0{,}4737*\log(HDP) + 0{,}0094*D(NUM(-1)) + 0{,}8651*\log(BV(-1)) + e_t$

S_{bj}	(0,47)	(0,031)	(0,062)	(0,0036)	(0,023)
t	(9,25)	(7,92)	(-7,59)	(2,59)	(36,67)

R2 = 0,996743
F – štatistika = 2448,385

[2] Lagrange multiplier tests
[3] Breusch-Godfrey Serial Correlation LM test

Odhady parametrov daných premenných sú vlastne bodovými koeficientmi elasticity. Ak vzrastie predchádzajúci príjem o jeden percentuálny bod, tak bankové vklady vzrastú v priemere o 0,25 percentuálneho bodu, za predpokladu ceteris paribus. Ak vzrastie výška bankových vkladov minulého obdobia o jeden percentuálny bod, tak výška bankových vkladov vzrastie v priemere o 0,86 percentuálneho bodu, za predpokladu ceteris paribus. Ak však vzrastie výška dosiahnutého hrubého domáceho produktu o jeden percentuálny bod, tak celková výška bankových vkladov nečakane klesne v priemere o 0,47 percentuálneho bodu, za predpokladu ceteris paribus, čo sa dá z časti vysvetliť posledným vývojom, keď hrubý domáci produkt rástol, no rast výšky bankových vkladov sa spomalil a niekedy až klesal. Ak vzrastie rozdiel nominálnych úrokových mier dvoch po sebe idúcich predchádzajúcich období o jednu jednotku, tak výška bankových vkladov vzrastie v priemere o 0,0094 percentuálneho bodu, za predpokladu ceteris paribus.

Všetky predpoklady väzieb medzi vysvetľujúcimi premennými (regresormi) a vysvetľovanou premennou sa až na jeden prípad potvrdili (okrem väzby BV a SAX, kde medzi danými premennými neexistuje skoro žiadna lineárna závislosť, pričom SAX je aj nevýznamnou premennou v danom modeli).

Nevýznamnosť niektorých premenných bola dosť prekvapujúca, no na druhej strane sa to dá vysvetliť nutnosťou držania peňazí na bežných bankových účtoch aj pri extrémnych hodnotách niektorých ukazovateľov, ktoré by mali mať podľa teórie silný, či už negatívny alebo pozitívny vplyv na výšku bankových vkladov.

Literatúra

HUŠEK, R. 1999. Ekonometrická analýza, Praha: Ekopress, 1999, ISBN 80-86119-19-X.

[b.a.] 2005 Sezónne očistený HDP a jeho zložky v mil. Sk stálych cien. Štatistický úrad SR. www.statistics.sk,

[b.a.] 2005 Tvorba a použitie dôchodkov v sektore domácností v mil. Sk bežných cien. Štatistický úrad SR. www.statistics.sk,

[b.a.] 2005. Vývoj priemerných úrokových mier korunových vkladov. Národná banka Slovenska. www.nbs.sk,

[b.a.] 2005. Vybrané makroekonomické ukazovatele. Národná banka Slovenska. www.nbs.sk

[b.a.] 2005. Analytické účty bankového sektora. NBS. www.nbs.sk,

[b.a.] 2005. História indexu SAX. Burza cenných papierov Bratislava. www.bcpb.sk

[b.a.] 2005. Archív kurzových lístkov. Národná banka Slovenska. www.nbs.sk,

LISÝ, J. a kol. 2000. Ekonómia. Bratislava, Edícia Ekonómia, tretie, prepracované a doplnené vydanie, 2000, ISBN 80-88715-81-4

BORZOVÁ, A.; MEDVEĎ, J. a kol. 1997. Úvod do teórie financií a meny. Banská Bystrica: Fakulta financií, 1997.

Adresa

Bc.Ján Beracko, 4. ročník
Ekonomická fakulta UMB
Tajovského 10
975 90 Banská Bystrica
jan_beracko@atlas.sk

Aplikácia Keranovho modelu akciových kurzov na index SAX

Martin Boďa

Abstrakt

The aim of this paper is to inquire into a possibility of applying Keran's stock price equation (as introduced in an article of his of January 1971) in the environment of the Slovak economy. The linear construction of the original model assumed that average investor behaviour may be explained by the adaptive expectations hypothesis, and that decisions are rendered in the light of real variables. The model being estimated on the latter assumption, however, leads to lags in explanatory variables of considerable length; which upon recognizing that that is not appropriate to the milieu of a *fledgling* stock market resulted in the necessity to modify the model. In this respect, real variables were substituted for nominal ones, and, further, the function form was slightly modified so as to produce acceptable length of lags. Both the models proved statistically significant, and it was shown that stock price (measured by the SAX index quarterly averages) is determined by the four factors assumed.

Úvod

V súčasnosti sú kapitálové trhy konvenčne považované za jeden z najdôležitejších atribútov rozvinutých ekonomík. Zmeny trhových podmienok sú dokumentované v hodnotách indexov kapitálového trhu, ktoré sú konštruované ako integrálny indikátor meniacej sa situácie. Je preto opodstatnená snaha finančných ekonómov popisovať vývoj trhových indexov a identifikovať determinanty, ktoré ich ovplyvňujú v najväčšej miere.

Jeden z možných prístupov, globálna analýza akciových trhov, akcentuje pôsobenie makroekonomických faktorov na pohyb akciových kurzov jednotlivých spoločností aj akciových indexov. V tomto kontexte je zámerom predkladaného príspevku preskúmať možnosť aplikácie globálneho modelu M. W. Kerana pre popísanie kurzotvorného mechanizmu slovenského akciového trhu sledovaného indexom SAX[1].

Pôvodný Keranov model

Keran (1971) nadviazal na st.-louiský model americkej ekonomiky z roku 1970[2] a zostavil rovnicu, v ktorej ceny akcií SP_t boli lineárne determinované 1. zmenami v reálnej peňažnej zásobe ΔM^*, 2. zmenami v reálnom produkte ekonomiky ΔX, 3. zmenami v cenovej hladine sledovanej cenovým indexom ΔP a 4. reálnymi ziskami korporácií E^*, čo produkovalo rovnicu zohľadňujúcu oneskorený vplyv zahrnutých premenných v tvare

$$SP_t = \beta_0 + \sum_{i=0}^{i=n1} \beta_{1i} \Delta M^*_{t-1} + \sum_{i=0}^{i=n2} \beta_{2i} \Delta X_{t-i} - \sum_{i=0}^{i=n3} \beta_{3i} \Delta P_{t-i} + \sum_{i=0}^{i=n4} \beta_{4i} E^*_{t-1} + \varepsilon_t , \qquad (1)$$

kde reálne veličiny M^* a E^* sú definované ako podiel nominálnej veličiny a cenového indexu $M^* = M/P$, resp. $E^* = E/P$. Následne po zavedení váh časových posunov je možné funkčný vzťah (1) vyjadriť v tvare

$$SP_t = \beta_0 + \beta_1 \sum_{i=0}^{i=n1} w_{1i} \Delta\left(\frac{M_{t-i}}{P_{t-i}}\right) + \beta_2 \sum_{i=0}^{i=n2} w_{2i} \Delta X_{t-i} + \beta_3 \sum_{i=0}^{i=n3} w_{3i} \Delta P_{t-i} + \beta_4 \sum_{i=0}^{i=n4} w_{4i} \left(\frac{E_{t-i}}{P_{t-i}}\right) + \varepsilon_t . \qquad (2)$$

---

[1] Index SAX je oficiálnym indexom Burzy cenných papierov Bratislava, a. s. Popisuje vývoj trhovej kapitalizácie súboru akcií zahrnutých do indexu a konštruuje sa ako cenový agregátny index v Paascheho tvare, t. j. $SAX_t = \sum p_t q_t / \sum p_{ref} q_t f$, kde p_t a p_{ref} sú ceny akcií k bežnému dňu (t) a k referenčnému dňu 14. 09. 1993 (ref), q_t je počet akcií k bežnému dňu (t) a f je korekčný faktor na zachovanie kontinuity pri zmene v štruktúre akcií zahrnutých do indexu.

[2] Bližšie pozri napr. ANDERSEN, LEONALL C., CARLSON, KEITH M.: *A Monetarist Model for Economic Stabilization.* In *Review*. The Federal Reserve Bank of St. Louis. Apríl 1970. S. 7 - 25.

Vymedzenie rovnice vzťahom (1), resp. (2) vychádza (explicitne) z monetaristickej hypotézy adaptívnych očakávaní a (implicitne) z tézy o rozhodovaní (priemerného) investora na základe reálnych, a nie nominálnych veličín.

Zdroje dát a metodika

Pri modelovaní boli použité kvartálne priemery indexu SAX za obdobie od Q3 1995 po Q2 2005 (48 údajov) skonštruované z denných hodnôt. Informácie o použitých dátach, ich zdrojoch a potrebných transformáciách ku kvartálnej frekvencii udáva tabuľka č. 1. Boli použité pôvodné dáta neočistené o periodickú zložku.

Ozn.	Popis	M. j.	Zdroj	Frekvencia dát	Adjustácia dát
SP	Hodnota akciového indexu SAX	body	*www.bcpb.sk*	denná (obchodné dni)	jednoduché priemerovanie
X	Reálny HDP, HDP v stálych cenách roku 1995 (ESA 1995) [3]	mld. Sk	*www.etrend.sk www.nbs.sk*	kvartálna	x
Y	Nominálny HDP, HDP v bežných cenách (ESA 1995) [3]	mld. Sk	*www.etrend.sk www.nbs.sk*	kvartálna	x
M	Peňažná zásoba (sledovaná agregátom M2 v bežných cenách)[4]	mld. Sk	*www.etrend.sk www.nbs.sk*	mesačná	chronologické priemerovanie
E	Zisky korporácií (sledované výsledkom hospodárenia nefinančných korporácií v bežných cenách)[5]	mld. Sk	*ŠÚ SR, databáza Slovstat*	ročná	pre jednotlivé kvartály boli bez úpravy použité ročné hodnoty
P	Index cenovej hladiny charakterizujúci vývoj spotrebiteľských cien oproti roku 1995 (k bázickému obdobiu Q2 1995 / Q3 1995)[6]	– (index)	mesačné zmeny spotrebiteľský cien *www.etrend.sk*	mesačná	kumulovanie mesačných zmien spotrebiteľských cien (Q2 1995 / Q3 1995 = 100 %)

Tabuľka č. 1

Údaje o ziskoch korporácií boli dostupné najdlhšie za nefinančné korporácie, aj to ročné dáta. Pre účely analýzy sa využili ročné hodnoty, ktoré bez adjustácie figurovali v štyroch kvartáloch príslušného roka.

So zreteľom na existenciu kolinearity medzi časovo posunutými premennými bola aplikovaná hypotéza Shirley Almonovej o polynomickom rozložení časových posunov a parametre $\beta_j w_{ij},\ i=1,2,\ldots,n_j,\ j=1,2,3,4$ v (2), resp. (3) boli aproximované polynómom r_j-teho stupňa

$$\beta_j w_{ij} = a_{(0)j} + a_{(1)j} i + a_{(2)j} i^2 + \ldots + a_{(rj)j} i^r \tag{3}$$

a následne bol vzťah (1) odhadovaný rovnicou

$$\begin{aligned} SP_t = \beta_0 &+ \left[a_{(0)1} \sum_{i=0}^{i=n1} \Delta M^*_{t-1} + \sum_{j=1}^{j=r1} a_{(j)1} \sum_{i=0}^{i=n1} i^j \Delta M^*_{t-1} \right] + \left[a_{(0)2} \sum_{i=0}^{i=n2} \Delta X_{t-i} + \sum_{j=1}^{j=r2} a_{(j)2} \sum_{i=0}^{i=n2} i^j \Delta X_{t-i} \right] + \\ &+ \left[a_{(0)3} \sum_{i=0}^{i=n3} \Delta P_{t-i} + \sum_{j=1}^{j=r3} a_{(j)3} \sum_{i=0}^{i=n3} i^j \Delta P_{t-i} \right] + \left[a_{(0)4} \sum_{i=0}^{i=n4} E^*_{t-1} + \sum_{j=1}^{j=r4} a_{(j)4} \sum_{i=0}^{i=n4} i^j E^*_{t-1} \right] + \varepsilon_t . \end{aligned} \tag{4}$$

Pre samotnú analýzu bol využitý ekonometrický softvér EViews. Pri zadávaní bolo potrebné špecifikovať okrem funkčného tvaru modelu i dĺžku oneskorenia vysvetľujúcich premenných (nebolo podmienkou, aby oneskorenie bolo rovnaké), stupeň aproximačného polynómu (volil sa 2, 3, max. 4) a obmedzenia polynomickej funkcie (preferovalo sa žiadne obmedzenie alebo obmedzenie polynómu na konci, t. j. na vystihnutie slabnutia efektu). Posledné tri nastavenia boli zadávané experimentálne a postupne sa prechádzalo k vyššej dĺžke oneskorenia, kým neboli dosiahnuté uspokojivé hodnoty indexu determinácie R^2, resp.

[3] Keran v konzistencii s vtedajšou preferenciou použil pre odhad produktu hrubý národný produkt (HNP). Teraz sa produkt konvenčne meria hrubým domácim produktom (HDP).

[4] Agregát M2 pozostáva z agregátu M1 a terminovaných účtov podnikov, obyvateľstva a poisťovní (kvázipeniaze). Pritom agregát M1 je zložený z obeživa mimo bánk a neterminovaných bankových vkladov.

[5] Do výkazníctva boli zahrnuté podniky s viac ako 20 zamestnancami.

[6] V pôvodnom modeli bol index cenovej hladiny zostavovaný na báze deflátora hrubého domáceho produktu.

adjustovaného indexu determinácie R^2_{adj}, Durbinovej-Watsonovej štatistiky DW, a dbalo sa na štatistickú významnosť odhadnutých parametrov modelu β_0 a $\beta_j w_{ij}$, $i = 1, 2, \ldots, n_j$, $j = 1, 2, 3, 4$. Výsledkom je model odhadnutý z 28 dát za obdobie od Q1 1998 po Q2 2005 a prezentovaný v schéme č. 1.

ODHAD PRE KRÁTKE OBDOBIE	$\text{est}SP_t = -269.8 - 55.67\sum_{i=0}^{i=14} w_{1i}\Delta M^*_{t-i} + 433.6\sum_{i=0}^{i=14} w_{2i}\Delta X_{t-i} - 2936\sum_{i=0}^{i=9} w_{3i}\Delta P_{t-i} - 0.009\sum_{i=0}^{i=14} w_{4i}E^*_{t-i}$
ODHAD PRE DLHÉ OBDOBIE	$\text{est}SP_t = -269.8 - 55.67\sum_{i=0}^{i=14} \Delta M^*_{t-i} + 433.6\sum_{i=0}^{i=14} \Delta X_{t-i} - 2936\sum_{i=0}^{i=9} \Delta P_{t-i} - 0.009\sum_{i=0}^{i=14} E^*_{t-i}$ (5)

		i	b_1w_{1i}	t_i	i	b_2w_{2i}	t_i	i	b_3w_{3i}	t_i	i	b_4w_{4i}	t_i
estβ_0	-269.8	$\Sigma\Delta M^*_i$	-55.67	-6.34	$\Sigma\Delta X_i$	433.6	5.00	$\Sigma\Delta P_i$	-2936	-2.21	ΣE^*_i	-0.00894	-3.71
t(est β_0)	-2.024												
R^2	0.985	**SSE**	14.31		**AIC**	8.446		**F**	109.7		**D-W**	1.825	
R^2_{ADJ}	0.976	**SE_{REGR}**	3480		**SC**	8.969		**p(F)**	0.000				

OBMEDZENIA: polynóm 4. stupňa pre ΔM^*, ΔX, E^* a polynóm 2. stupňa pre ΔP;
$\Delta M^*_1 = 0$, $\Delta M^*_{-(n1+1)} = 0$; $\Delta X_1 = 0$, $\Delta X_{-(n2+1)} = 0$; $\Delta P_1 = 0$, $\Delta P_{-(n3+1)} = 0$; $E^*_1 = 0$, $E^*_{-(n4+1)} = 0$

Schéma č. 1

Odhadnutý funkčný vzťah (5) je zaťažený vysokou dĺžkou oneskorenia (v prípade troch premenných dokonca tri a pol roka), čo prichodí s ohľadom na podmienky dubiózne. Súbežne nie sú rešpektované požiadavky ekonomickej teórie na spôsob, akým vysvetľujúce premenné determinujú cenu akcií: nárast peňažnej zásoby má stimulovať rast cien akcií a vyššie zisky korporácií tiež majú mať pozitívny kurzotvorný účinok.

Modifikovaný Keranov model (vzhľadom na podmienky slovenskej ekonomiky)

Imanentným nedostatkom Keranovho modelu je skutočnosť, že bol vyvinutý pre chronicky inflačnú americkú ekonomiku 70. rokov. Na účely aplikácie pre podmienky slovenskej republiky je potrebné zohľadniť odlišné spoločenské, hospodárske a nové podmienky.

Ak odmietneme predpoklad o rozhodovaní sa na základe reálnych veličín a zoberieme do úvahy, že investor (akokoľvek racionálny) nedokáže rozlíšiť účinok zmien v cenovej hladine na reálne veličiny (hoci si nevyhnutne uvedomuje meniacu sa cenovú hladinu) a že rozhodnutia prijíma na základe nominálnych veličín, v prostredí ktorých sa nachádza, je potrebné pôvodný model modifikovať. Prejaví sa to v používaní nominálneho produktu ekonomiky (Y) a peňažná zásoba (M) a zisky korporácií (E) nebudú pomerované indexom cenovej hladiny. Na základe tejto požiadavky slovenského prostredia a po zavedení procesov náhodnej zložky AR(1), AR(2) a MA(1) možno regresnú úlohu formulovať vzťahmi

$$SP_t = \beta_0 + \beta_1 \sum_{i=0}^{i=n1} w_{1i}\Delta M_{t-i} + \beta_2 \sum_{i=0}^{i=n2} w_{2i}Y_{t-i} + \beta_3 \sum_{i=0}^{i=n3} w_{3i}P_{t-i} + \beta_4 \sum_{i=0}^{i=n4} w_{4i}E_{t-i} + \varepsilon_t\,, \qquad (6a)$$

$$\varepsilon_t = \phi_1\varepsilon_{t-1} + \phi_2\varepsilon_{t-2} + \theta_1\eta_{t-1} + \eta_t\,, \qquad (6b)$$

kde ϕ_i je koeficient autokorelácie náhodných zložiek modelu *i*-teho rádu a η_t je inovačný člen. Takto špecifikovaný problém vedie k odhadu s výstupmi znázornenými v schéme č. 2. (Pre odhad poslúžilo 38 pozorovaní za obdobie od Q1 1996 po Q2 2005.) Model v snahe eliminovať potenciálne problémy s autokoreláciou i heteroskedasticitou bola odhadnutý zovšeobecnenou metódou najmenších štvorcov za použitia Neweyho-Westovho kovariančného estimátora.

Absencia heteroskedasticity bola overovaná Whiteovým testom, testom založeným na Spearmanovom koeficiente poradovej korelácie, ako aj Bartlettovým testom a Cochranovým testom pri porovnávaní rozptylov čiastkových súborov rezíduí. Všetky testy na všetkých hladinách významnosti nepotvrdili prítomnosť heteroskedasticity. Odhadnutý model (9) bol

zbavený štatisticky významnej autokorelácie rezíduí, čo potvrdil korelogram rezíduí aj štvorcových rezíduí a Ljungova-Boxova Q štatistika. Rovnako testy autokorelácie rezíduí prvého rádu založené na Durbinovej-Watsonovej štatistike a von Neumannovom pomere taktiež neviedli k potvrdeniu jej existencie. Rezíduá sú normálne rozdelené, čo na všetkých bežných hladinách významnosti potvrdili o. i. Jarqueov-Berov test a Shapirov-Wilkov test. I keď model disponuje určitými štatistickými kvalitami, problémom zostáva jeho stabilita: testy štrukturálnych zmien (Chowov test a Ramseyov RESET test) indikujú jeho nestabilitu.

ODHAD PRE KRÁTKE OBDOBIE

$$\text{est}SP_t = 531.1 + 11.2\sum_{i=0}^{i=3} w_{1i}\Delta M_{t-i} + 7.9\sum_{i=0}^{i=4} w_{2i}Y_{t-i} - 1633\sum_{i=0}^{i=3} w_{3i}P_{t-i} + 0.002\sum_{i=0}^{i=2} w_{4i}E_{t-i} - 0.96e_{t-1} - 0.82e_{t-2} - 1.00\eta_{t-1}$$

ODHAD PRE DLHÉ OBDOBIE

$$\text{est}SP_t = 531.1 + 11.2\sum_{i=0}^{i=3} \Delta M_{t-i} + 7.9\sum_{i=0}^{i=4} Y_{t-i} - 1633\sum_{i=0}^{i=3} P_{t-i} + 0.002\sum_{i=0}^{i=2} E_{t-i} - 0.96e_{t-1} - 0.82e_{t-2} - 1.00\eta_{t-1} \quad (9)$$

		i	b_1w_{1i}	t_i	i	b_2w_{2i}	t_i	i	b_3w_{3i}	t_i	i	b_4w_{4i}	t_i
estβ_0	531.10	ΔM_0	2.337	7.01	Y_0	1.235	3.60	P_0	-318.8	-4.94	E_0	0.00026	1.94
t(estβ_0)	51.027	ΔM_{-1}	3.873	9.96	Y_{-1}	1.891	6.73	P_{-1}	-486.1	-11.2	E_{-1}	0.00059	32.5
estβ_1	11.213	ΔM_{-2}	3.316	8.66	Y_{-2}	2.032	16.8	P_{-2}	-494.0	-15.7	E_{-2}	0.00064	4.47
estβ_2	7.9007	ΔM_{-3}	1.684	3.83	Y_{-3}	1.721	5.47	P_{-3}	-334.5	-6.11	ΣE^*_i	0.00149	32.5
estβ_3	-1633	$\Sigma\Delta M_i$	11.21	10.8	Y_{-4}	1.022	2.80	ΣP_i	-1633	-20.2			
estβ_4	0.0015				ΣY_i	7.901	16.8						
estϕ_1	0.9618												
t(estϕ_1)	10.575				R^2	0.978		SE_{REG}	14.32		D-W	2.1111	
estϕ_2	-0.815	estθ_1	-0.997		R^2_{ADJ}	0.968		AIC	8.427		F	96.4839	
t(estϕ_2)	-7.673	t(estθ_1)	-4.073		SSE	5129		SC	8.987		p(F)	0.000000	

OBMEDZENIA: polynóm 3. stupňa pre všetky premenné; $\Delta M_1 \neq 0$, $\Delta M_{-(n1+1)} = 0$; $Y_1 = 0$, $Y_{-(n2+1)} = 0$; $P_1 = 0$, $P_{-(n3+1)} = 0$; $E_1 = 0$, $E_{-(n4+1)} = 0$

Schéma č. 2

S prihliadnutím na cieľ práce, ktorým bolo preskúmať kurzotvorné pôsobenie štyroch faktorov na akciový index SAX, však možno uviesť, že model je v týchto reláciách postačujúci. Bolo ukázané, že priemerný investor je pri rozhodovaní ovplyvňovaný nielen súčasnými nominálnymi veličinami, ale pamätá si i stav týchto veličín za obdobie pol až jeden rok dozadu. Zmeny týchto veličín ovplyvňujú hodnoty indexu SAX smerom, ktorý im prisudzuje súčasný prúd finančnej ekonómie (či už v dlhom, alebo v krátkom období). Ak vezmeme do úvahy, že množstvo peňazí v obehu (resp. peňažná zásoba) a inflačný rast cien sú výrazom monetárnej politiky štátu, možno oneskorenie tri kvartálne obdobia vysvetľovať ako rozložený časový posun efektov transmisného mechanizmu politiky Národnej banky Slovenska[7]. Výsledky vytvárajú tiež širší priestor pre diskusiu o význame zaradenia ziskov korporácií po zdanení do modelu, nakoľko ich koeficienty sú blízke jednej (v porovnaní s hrubým domácim produktom alebo zmenami v peňažnej zásobe, ktoré sú tiež v rovnakej vykazovacej jednotke mld. Sk), napriek tomu sa ukázal ich pozitívny vplyv na hodnoty indexu SAX.

Použitá literatúra

[1.] GREENE, WILLIAM H.1997. *Econometric Analysis.* Tretie vydanie London Prentice Hall 1997. 1075 s. ISBN 0-13-724659-5.

[2.] HUŠEK, ROMAN 1999. *Ekonometrická analýza.* Prvé vydanie Praha. Ekopress 1999. 304 s. ISBN 80-86119-19-X.

[3.] KERAN, MICHAEL W. 1971. *Expectations, Money, and the Stock Market.* In *Review*. The Federal Reserve Bank of St. Louis. Január 1971. S. 16 - 31.

[7] Podľa [6.] vo *vyspelých* ekonomikách je bežné oneskorenie šiestich mesiacov.

[4.] MUSÍLEK, P. 2002. *Trhy cenných papírů.* Praha. Ekopress 2002. 460 s. ISBN 80-86119-55-6.
[5.] *SAX Index.* Burza cenných papierov Bratislava, a. s. [Cit. k 01. 11. 2005] Dostupné na World Wibe Web: <http://www.bsse.sk/Obchodovanie /Indexy/IndexSAX.aspx#>.
[6.] ZIMKOVÁ, E. 2002. *Menová politika.* Prvé vydanie. UMB v Banskej Bystrici, Fakulta financií 2002. 83 s. ISBN 80-8055744-6.

Adresa
Bc. Martin Boďa, 4. ročník
Ekonomická fakulta UMB
Tajovského 10
975 90 Banská Bystrica
ma_bo@azet.sk

Analýza dopravnej nehodovosti v okresoch Slovenska

Karol Frišták

Súhrn
Cieľom práce bolo analyzovať nehodovosť na cestách v okresoch Slovenska v roku 2003.na základe štyroch relatívnych ukazovateľov metódami viacrozmernej analýzy.
Kľúčové slová
Metóda poradí, analýza hlavných komponentov, zhluková analýza.

Summary
The aim of this work was to analyze an accident frequency at the roads of the districts within Slovakia in the year 2003 on the base of four relative indicators by applying of multidimensional methods.
Key words
Simple method of ranking, Principal Component Analysis, Cluster Analysis.

1. Dáta

Dáta o nehodovosti v okresoch Slovenska za rok 2003 sme získali z internetovej stránky ministerstva vnútra SR (www.minv.sk). Náš dátový súbor obsahuje 5 premenných a 77 štatistických jednotiek. Tieto jednotky sú vlastne okresy na Slovensku[1]. Celý dátový súbor poskytuje informácie o nehodovosti na cestách okresov SR nielen z pohľadu celkového počtu nehôd, ale aj z pohľadu závažnosti nehôd v daných okresoch. Sú tu uvedené údaje o počtoch usmrtených, ťažko a ľahko zranených.
Pre viacrozmernú analýzu nehodovosti za rok 2003 sme z pôvodných absolútnych premenných vypočítali nasledujúce relatívne ukazovatele (ďalej len ukazovatele):

- **Neh_1000oby03** - *Počet nehôd na 1000 obyvateľov* = celkový počet nehôd / počet obyvateľov x 1000
- **Usm_z_celk03** - *Počet usmrtených osôb na 1000 dopravných nehôd* = počet usmrtených osôb / celkový počet nehôd x 1000
- **Taz_z_celk03** - *Počet ťažko ranených na 1000 dopravných nehôd* = počet ťažko ranených osôb / celkový počet nehôd x 1000
- **Lah_z_celk03** - *Počet ľahko ranených na 1000 dopravných nehôd* = počet ľahko ranených osôb / celkový počet nehôd x 1000

2. Cieľ

Cieľom našej analýzy je zhodnotiť úroveň nehodovosti v jednotlivých okresov podľa vytvorených relatívnych ukazovateľov. Ide o viacrozmernú analýzu a na dosiahnutie cieľa použijeme nasledovné metódy:

- jednoduchú metódu poradí,
- analýzu hlavných komponentov,
- zhlukovú analýzu.

[1] Ako vieme na Slovensku je 79 okresov, ale na stránke www.minv.sk sú uvedené dáta len za 77 okresov.

3. Jednoduché viacrozmerné porovnávanie na základe poradia

Nahradením hodnôt ukazovateľov hodnotami ich poradí, sme jednotlivým okresom priradili poradia od najlepších (1. poradie) po najhoršie (77. poradie). Z poradí za všetky ukazovatele pre okresy sme vytvorili priemerné poradie a okresy sme zoradili (viď tabuľka č. 1).
Ako vidieť z tejto tabuľky na celkové poradie (premenná rank_priem03) naozaj vplýva nielen počet nehôd na 1000 obyvateľov, ale aj závažnosť nehôd, či už počet usmrtených, ťažko i ľahko zranených.
Skupina 5 najlepších okresov: Na čele tejto skupiny je okres Medzilaborce s najmenším počtom dopravných nehôd na 1000 obyvateľov a celkovo dobrými ostatnými ukazovateľmi. Na ďalších štyroch miestach sú prekvapujúco bratislavské okresy, ktoré sú najhoršie v počte dopravných nehôd na 1000 obyvateľov, ale v ukazovateľoch závažnosti sú medzi najlepšími.
Skupina 5 najhorších okresov: na týchto posledných miestach sa umiestnili okresy nie tak zlé v počte dopravných nehôd na 1000 obyvateľov, ale skôr v závažnosti dopravných nehôd. Tu sú okresy s najvyššími počtami usmrtených, ťažko či ľahko zranených osôb.

Tabuľka č. 1: Časť tabuľky priemerných poradí (5 najlepších a 5 najhorších okresov)

Okres	Neh_1000oby03	usm_z_celk03	taz_z_celk03	lah_z_celk03	rank_Neh_1000oby03	rank_usm_z_celk03	rank_taz_z_celk03	rank_lah_z_celk03	**rank_priem03**
Medzilaborce	3.55	0	22.22	155.55	1	1.5	9	19	**7.625**
Bratislava I	53.12	0.42	9.24	39.49	77	3	1	1	**20.5**
Bratislava V	18.84	1.31	10.50	77.02	71	4	2	6	**20.75**
Bratislava IV	17.10	4.39	13.19	59.04	69	12	4	2	**21.75**
Bratislava II	28.76	1.92	13.50	60.75	75	5	5	3	**22**
atď.									
Gelnica	5.12	75.94	63.29	373.41	4	77	66	77	**56**
Michalovce	9.44	18.42	64.01	219.20	44	58	67	56	**56.25**
Zlaté Moravce	7.47	33.74	58.28	285.27	23	73	62	73	**57.75**
Sobrance	6.56	38.46	76.92	269.23	13	75	75	68	**57.75**
Lučenec	8.26	29.90	69.76	275.74	29	70	72	69	**60**

Záver: Okres Medzilaborce má najmenší počet obyvateľov a je celkovo najmenší. Nie je v ňom veľmi rozšírená infraštruktúra a preto je najlepší z pohľadu nehodovosti. Bratislavské okresy sa tiež zaradili medzi najlepších, majú však najväčší počet dopravných nehôd, ktoré ale nie sú závažné. To znamená v Bratislave vznikajú dopravné nehody tzv. "ťukačky".
Okresy, ktoré sa zaradili na chvost tabuľky sa nachádzajú väčšinou na hlavnom ťahu Bratislava - Košice. Tu sa nenachádzajú rýchlostné komunikácie a cesty sú preplnené vozidlami. Cesty sú tu v zlom stave a vozidlá nespĺňajú základné technické požiadavky, čo sa odráža na celkovej nehodovosti.

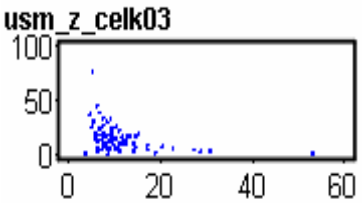
4. Závislosť ukazovateľov

Medzi ukazovateľmi existuje štatisticky významná korelácia, čiže ukazovatele sú navzájom závislé ako je možné vidieť z korelačnej matice (Tabuľka č. 2).

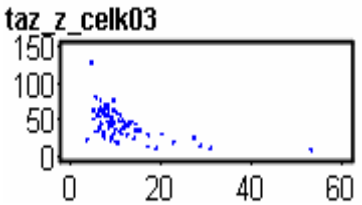
***Tabuľka č. 2*:** Pearsonová korelačná matica

	Neh_1000oby03	**usm_z_celk03**	**taz_z_celk03**	**lah_z_celk03**
Neh_1000oby03	1.00000	-0.38647 0.0005	-0.54695 <.0001	-0.66147 <.0001
usm_z_celk03	-0.38647 0.0005	1.00000	0.42037 0.0001	0.58682 <.0001
taz_z_celk03	-0.54695 <.0001	0.42037 0.0001	1.00000	0.61192 <.0001
lah_z_celk03	-0.66147 <.0001	0.58682 <.0001	0.61192 <.0001	1.00000

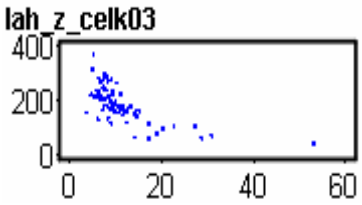
Z grafov 1 až 3 je zrejmé, že s počtom dopravných nehôd v okresoch klesá ich závažnosť, ide teda o nepriamu závislosť.



Graf1: Závislosť usmrtených na nehodách



***Graf2*:** Závislosť ťažko ranených na nehodách



***Graf3*:** Závislosť ľahko ranených na nehodách

5. Metóda hlavných komponentov (PCA)

Ďalej sme použili metódu hlavných komponentov a to jednak na identifikáciu okresov s extrémnymi hodnotami jednotlivých ukazovateľov a jednak na vytvorenie nezávislých premenných, ktoré neskôr použijeme v zhlukovej analýze.
Keďže sme zistili, že analyzované ukazovatele sú závislé, vytvoríme si hlavné komponenty. Tieto komponenty budú lineárnou kombináciou premenných a budú navzájom nezávislé. Pri výpočte hlavných komponentov použijeme korelačnú maticu (viď tabuľka č.2), pretože premenné sú namerané v rôznych merných jednotkách.

***Tabuľka č. 3*:** Vlastné čísla hlavných komponentov

Eigenvalues of the Correlation Matrix				
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	2.61939791	1.97617660	0.6548	0.6548
2	0.64322131	0.18422367	0.1608	0.8157
3	0.45899764	0.18061449	0.1147	0.9304
4	0.27838315		0.0696	1.0000

Z tabuľky č. 3 vyplýva, že 1. hlavný komponent vysvetľuje až 65,5% variability dát a 2. hlavný komponent vystihuje už len 16%. Zvyšné dva komponenty spolu vysvetľujú len 18,5% variability, čo môžeme považovať za zanedbateľné.

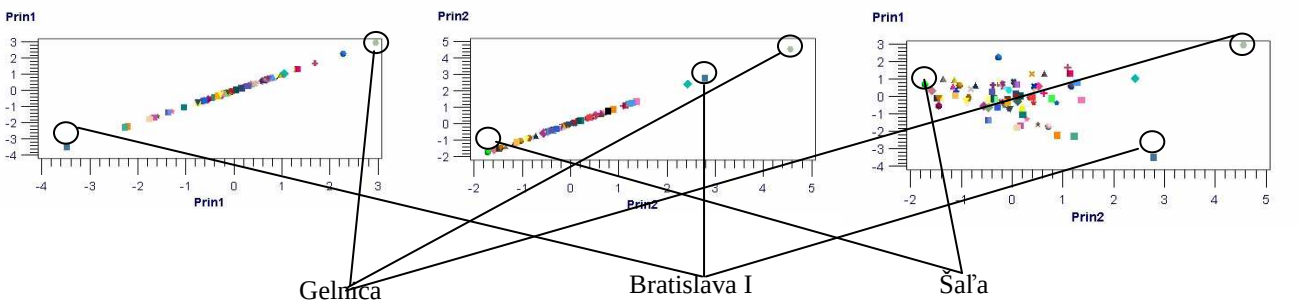
***Tabuľka č. 4*:** Váhy premenných v komponentoch PRIN1 a PRIN2

	Prin1	**Prin2**
Neh_1000oby03	-.499604	0.457617
usm_z_celk03	0.447852	0.831163
taz_z_celk03	0.493964	-.315088
lah_z_celk03	0.553013	0.021756

Ako vidieť z tabuľky 4, prvý hlavný komponent Prin1 je tvorený rovnakou váhou všetkými pôvodnými premennými (váhy sú okolo 0,5). Ukazovatele o počet usmrtených, ťažko a ľahko zranení majú však kladné znamienko, kým počet nehôd na 1000 obyvateľov má záporné znamienko. V 2. hlavnom komponente Prin2 má najvyššiu váhu ukazovateľ o počet usmrtených.
Na grafe 4 sú jednotlivé okresy zobrazené na priemete hlavných komponentov Prin1 a Prin2. Na prvom obrázku sú zobrazené na x-ovej aj

y-ovej osy hodnoty prvého hlavného komponentu. Ako extrémne hodnoty (outliers) sa prejavili okresy Gelnica, ktorá má veľmi nízky počet dopravných nehôd ale extrémne vysoký podiel usmrtených a Bratislava I, ktorá má veľmi vysoký počet dopravných nehôd, ale nie závažných.

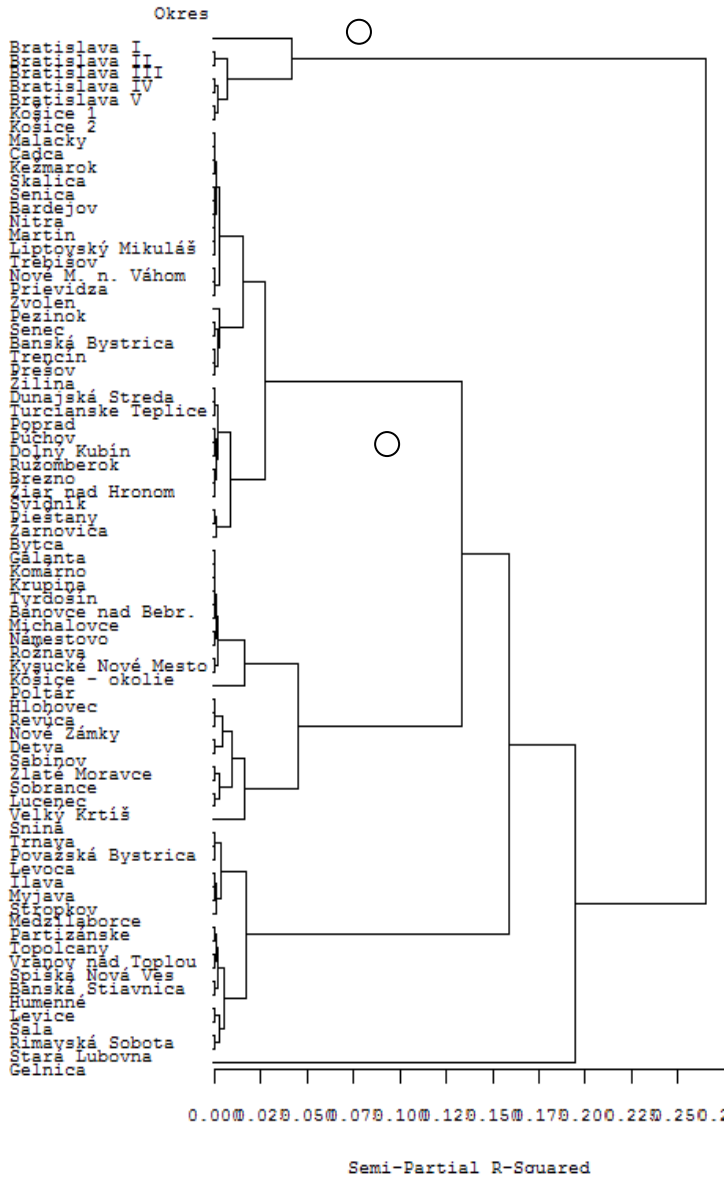
Graf 4: Zobrazenie okresov na priemete hlavných komponentov PRIN1 a PRIN2



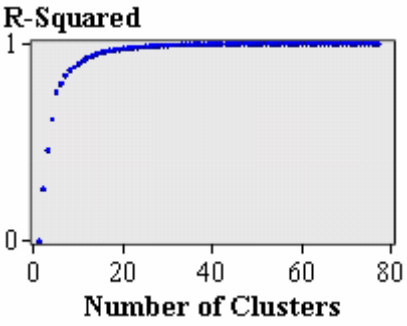
6. Zhluková analýza

Zhluková analýza patrí medzi klasifikačné viacrozmerné metódy, ktorá štatistické jednotky usporiada do zhlukov s podobnými vlastnosťami ukazovateľov.
V našej analýze sme použili hierarchické zhlukovanie Wardovou metódou. Výsledky môžeme vidieť na dendograme (viď obrázok 1).

Obrázok 1:: Dendrogram



Graf č 5: Koeficient determinácie



Z grafu 5 je zrejmé, že optimálny počet zhlukov by mohol byť medzi 7 až 15, pretože semiparciálne koeficienty derminácie (R-Squared) sa pre 7 a viac zhlukov blížia k číslu 1 a nie sú veľmi rozdielne. Zvolili sme 7 výsledných zhlukov aj na základe dendrogramu. Jednotlivé zhluky obsahujú okresy s podobnými vlastnosťami ohľadne nehodovosti. Obsah zhlukov je nasledovný:

Zhluk č. 1: Malacky, Pezinok, Senec, Dunajská Streda, Piešťany, Senica, Skalica, Trenčín, Nové M. n. Váhom, Prievidza, Púchov, Nitra, Žilina, Bytča, Čadca, Dolný Kubín, Liptovský Mikuláš, Martin, Ružomberok, Turčianske Teplice,

Banská Bystrica, Brezno, Zvolen, Žarnovica, Žiar nad Hronom, Prešov, Bardejov, Kežmarok, Poprad, Svidník, Trebišov.
Zhluk č. 2: Galanta, Bánovce nad Bebravou, Komárno Kysucké, Nové Mesto Námestovo, Tvrdošín, Krupina, Poltár, Košice – okolie, Michalovce Rožňava.

Zhluk č. 3: Trnava, Ilava, Myjava, Partizánske, Považská Bystrica, Levice, Šaľa, Topoľčany, Banská Štiavnica, Rimavská Sobota, Humenné, Levoča, Medzilaborce, Stará Ľubovňa, Stropkov, Vranov nad Topľou, Spišská Nová Ves.

Zhluk č. 4: Hlohovec, Nové Zámky, Zlaté Moravce, Detva, Lučenec, Revúca, Veľký Krtíš, Sabinov, Snina, Sobrance.
Zhluk č. 5: Bratislava II, Bratislava III, Bratislava IV, Bratislava V, Košice 1, Košice 2.
Zhluk č. 6: Bratislava I.
Zhluk č. 7: Gelnica.

7. Záver

Viacrozmernými metódami sme dosiahli rozdelenie celého súboru okresov Slovenska do 7 skupín (zhlukov). Môžeme povedať, že metóda poradí, metóda PCA a zhluková analýza nám pomohli rozanalyzovať súbor okresov SR z hľadiska nehodovosti. Metódy neprinášajú úplne identické výsledky.
Analyzovali sme aj situáciu v nehodovosti v za rok 2004. Výsledky však boli podobné a preto ich v tomto príspevku neuvádzame.

Literatúra

[1] SHARMA, S.: Applied Multivariate Techniques. John Wiley & Sons, Inc. 1996.
[2] STANKOVIČOVÁ, I.: Viacrozmerná analýza rentability poisťovní SR pomocou Enterprise Guide. In: 10. medzinárodný seminár Výpočtová štatistika. Bratislava: SŠDS 2001. ISBN 80-88946-14-X.
[3] VOJTKOVÁ, M.: Zhluková analýza a jej aplikácia na priemyselné podniky SR z hľadiska efektívnosti. FernStat 2004. Slovenská konferencia aplikovanej štatistiky. Zborník príspevkov. Tajov, s. 72-77, Bratislava: SŠDS, 2004. ISBN 80-88946-34-4
[4] SAS OnLine Documentation: http://support.sas.com/91doc/docMainpage.jsp

Kontakt:
Karol Frišták, Fakulta managementu UK, 3 ročník
Podjavorinskej 23, Nitra 94901

karol.fristak@st.fm.uniba.sk

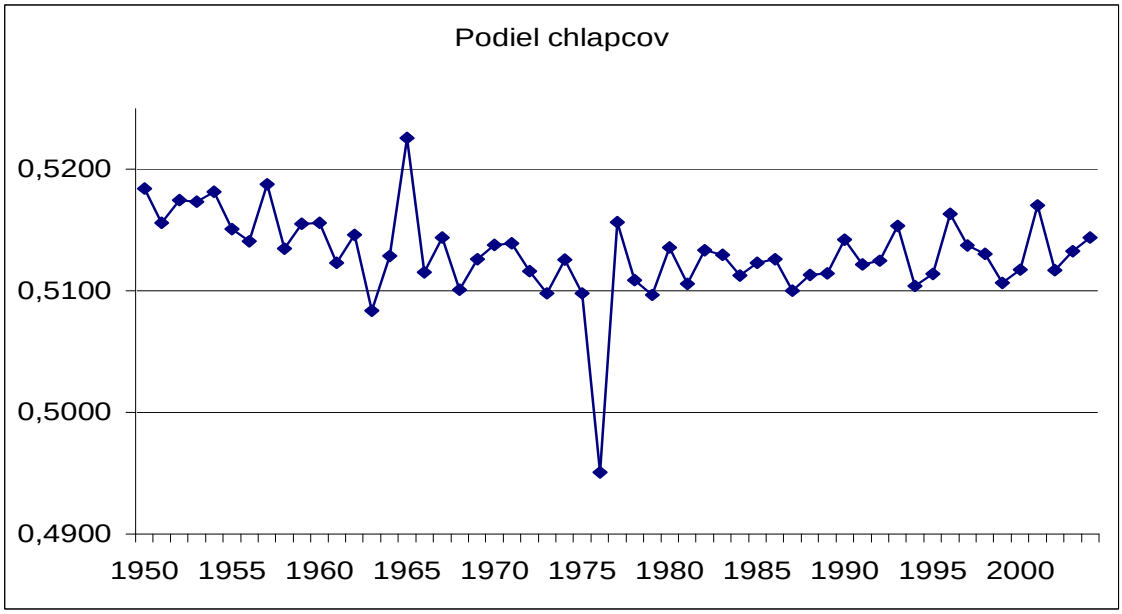
Analýza vývoja podielu živonarodených v SR podľa pohlavia v rokoch 1950 až 2004

Andrej Gerboc

Súbor, ktorý som analyzoval v mojej práci obsahuje údaje za 55 rokov o počte živonarodených detí celkom, ďalej chlapcov a dievčat na Slovensku v rokoch 1950 až 2004. Údaje, ktoré sú uvedené v prílohe, som získal so stránky Výskumného demografického centra:

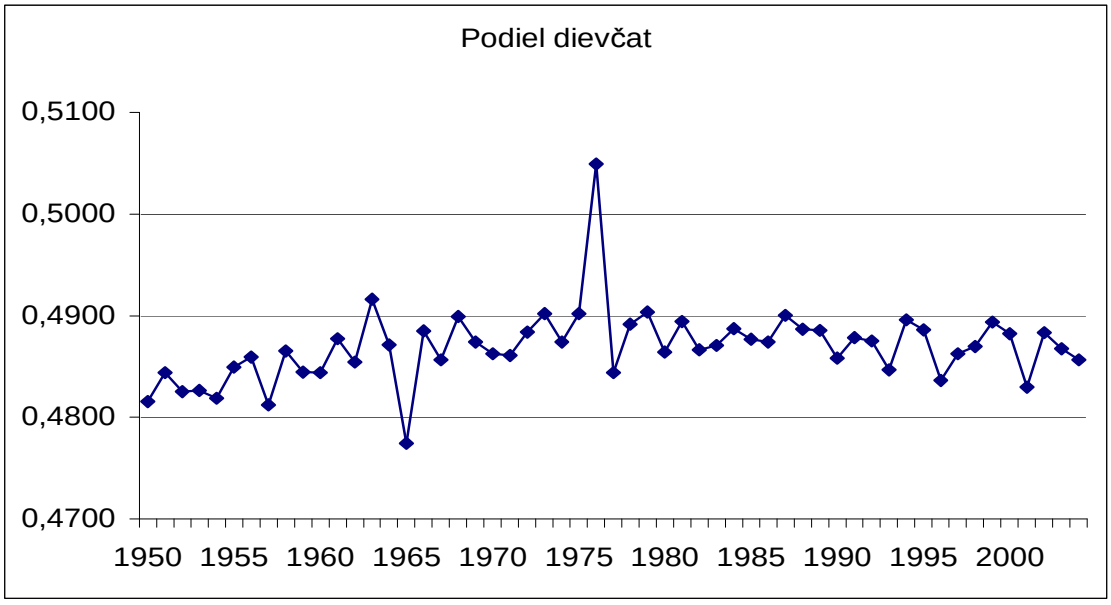
http://www.infostat.sk/slovakpopin/data/maindata.htm

Čo sa týka pomeru narodených chlapcov, maximum sa dosiahlo v roku 1965 a to 52,26% a minimum v roku 1976 - 49,51%. Priemerne dosiahol tento pomer hodnotu 0,513 čo znamená, že na 1000 živonarodených detí pripadá 513 chlapcov. Pomer dievčat bol najvyšší v roku 1976 – 50,49% a najnižší v 1965 a to 47,74% a v priemere pripadalo na 1000 živonarodených detí 487 dievčat. V literatúre sa zvykne tento pomer udávať číslom 0,486 – teda 486 dievčat na 1000 narodených detí a z mojej analýzy sa to potvrdilo takmer presne, rozdiel predstavuje o 1 narodené dievča na 1000 detí viacej.
Celkovo najviac detí sa narodilo v roku 1952 a to 100 824 a najmenej v roku 2002 - 50 841. V priemere sa za toto obdobie narodilo 83 037 detí, medián predstavoval 87 138 detí a smerodajná odchýlka bola 15 376 detí. Medzi chlapcami sa maximum dosiahlo v roku 1952 a to 52 171 a minimum v roku 2002 - 26 015. Priemerne sa narodilo 42 603 chlapcov. Medián bol 44 559 a smerodajná odchýlka 7897 chlapcov. Dievčat sa narodilo najviac v roku 1976 – 50 398 a najmenej 24 697 v roku 2001, priemer bol 40 433 dievčat. Medián má hodnotu 42 471 dievčat a smerodajná odchýlka je 7494 dievčat.
Čo sa týka vývoja pomeru narodených chlapcov z grafu (Obr. 1) je vidieť mierny pokles tohto podielu od roku 1950 do roku 1976, keď tento podiel dosiahol svoje minimum (49,51%), pričom výnimkou bol rok 1965, v ktorom dosiahol pomer chlapcov maximálnu hodnotu (52,26%). V rokoch 1978 – 1989 bol tento pomer približne rovnaký a od roku 1989 po 2004 je vidieť mierny nárast tohto podielu, ktorý dosiahol v tomto roku hodnotu 51,44 % t.j. 514 chlapcov na 1000 živonarodených detí.



Obr. 1 – Graf podielu chlapcov na živonarodených deťoch

Z grafu podielu dievčat (Obr. 2) je naopak vidieť nárast tejto hodnoty od roku 1950 do roku 1976, kedy podiel dosiahol maximálnu hodnotu (50,49%), s výnimkou v roku 1965, kedy bol tento pomer najmenší (47,74%). V rokoch 1978 až 1992 sa veľmi nemenil. Po roku 1994 (48,96%) je vidieť pokles podielu dievčat v rokoch 1995 a 1996 (48,36%) a následný nárast v rokoch 1997, 1998 a 1999 kedy dosiahol hodnotu 48,94% a opätovný pokles v rokoch 2000 a 2001. V roku 2002 zaznamenal pol percentný rast, aby v ďalších 2 rokoch 2003, 2004 opäť klesol a v minulom roku dosiahol hodnotu 48,56 %, teda 486 dievčat na 1000 živonarodených detí.
Z grafu je tiež vidieť, po odmyslení si minimálnej a maximálnej hodnoty, jeho parabolický priebeh. Zaujímavé by bolo tiež zistiť príčiny výrazných extrémnych hodnôt podielu dievčat v rokoch 1965 (minimum 47,74%) a 1976 (maximum 50,49%).



Obr. 2 – Graf podielu dievčat na živonarodených deťoch

Záverom mojej práce by som zhrnul, že podiel dievčat má po roku 1999 klesajúcu tendenciu, čo by mohlo v budúcnosti spoločne s celkovým poklesom počtu narodených detí mať negatívny vplyv na reprodukciu obyvateľstva SR
a tiež spôsobovať problémy so starnutím populácie.

Príloha 1 – vstupné údaje

Rok	Živonarodené deti	chlapci	dievčatá	podiel chlapcov	podiel dievcat
1950	99721	51699	48022	0,518	0,482
1951	100663	51903	48760	0,516	0,484
1952	100824	52171	48653	0,517	0,483
1953	99124	51281	47843	0,517	0,483
1954	98310	50939	47371	0,518	0,482
1955	99305	51149	48156	0,515	0,485
1956	99467	51135	48332	0,514	0,486
1957	97311	50481	46830	0,519	0,481
1958	93272	47894	45378	0,513	0,487
1959	87991	45362	42629	0,516	0,484
1960	88412	45584	42828	0,516	0,484
1961	87359	44753	42606	0,512	0,488

1962	83899	43174	40725	0,515	0,485
1963	87158	44308	42850	0,508	0,492
1964	86878	44559	42319	0,513	0,487
1965	84257	44031	40226	0,523	0,477
1966	81453	41664	39789	0,512	0,488
1967	77537	39883	37654	0,514	0,486
1968	76370	38955	37415	0,510	0,490
1969	79769	40890	38879	0,513	0,487
1970	80666	41443	39223	0,514	0,486
1971	83062	42687	40375	0,514	0,486
1972	87794	44918	42876	0,512	0,488
1973	92953	47387	45566	0,510	0,490
1974	97585	50020	47565	0,513	0,487
1975	97649	49783	47866	0,510	0,490
1976	99814	49416	50398	0,495	0,505
1977	99533	51322	48211	0,516	0,484
1978	100193	51185	49008	0,511	0,489
1979	100240	51087	49153	0,510	0,490
1980	95100	48842	46258	0,514	0,486
1981	93290	47631	45659	0,511	0,489
1982	92618	47545	45073	0,513	0,487
1983	92053	47218	44835	0,513	0,487
1984	90843	46446	44397	0,511	0,489
1985	90155	46189	43966	0,512	0,488
1986	87138	44667	42471	0,513	0,487
1987	84006	42843	41163	0,510	0,490
1988	83242	42563	40679	0,511	0,489
1989	80116	40976	39140	0,511	0,489
1990	79989	41130	38859	0,514	0,486
1991	78569	40241	38328	0,512	0,488
1992	74640	38253	36387	0,513	0,488
1993	73256	37752	35504	0,515	0,485
1994	66370	33875	32495	0,510	0,490
1995	61427	31415	30012	0,511	0,489
1996	60123	31045	29078	0,516	0,484
1997	59111	30368	28743	0,514	0,486
1998	57582	29543	28039	0,513	0,487
1999	56223	28710	27513	0,511	0,489
2000	55151	28224	26927	0,512	0,488
2001	51136	26439	24697	0,517	0,483
2002	50841	26015	24826	0,512	0,488
2003	51713	26543	25170	0,513	0,487
2004	53747	27646	26101	0,514	0,486

Príloha 2 – základný rozbor

spolu		chlapci		dievčatá	
počet	55	počet	55	počet	55
minimum	50841	minimum	26015	minimum	24697
maximum	100824	maximum	52171	maximum	50398
priemer	83037	priemer	42603	priemer	40433
medián	87138	medián	44559	medián	42471
stdev	15376,4	stdev	7897,72	stdev	7494,13

podiel chlapcov		podiel dievčat	
počet	55	počet	55
minimum	0,495	minimum	0,477
maximum	0,523	maximum	0,505
priemer	0,513	priemer	0,487
medián	0,513	medián	0,487

Použitá literatúra:
1.) Chajdiak, J.: Základy hospodárskej štatistiky
2.) Chajdiak, J.: Štatistické úlohy a ich riešenie v Exceli
3.) Jurčová, D.: Slovník demografických pojmov

Použitý software: Microsoft Excel a Mikrosoft Word

Kontakt na autora: gerboc.andrej@post.sk
Andrej Gerboc
FHI EU, 5. ročník, AŠ

Religiózna štruktúra obyvateľstva Slovenska – pokus o zachytenie prirodzeného pohybu piatich najpočetnejších cirkví na Slovensku

Viera Hluchá

Úvod

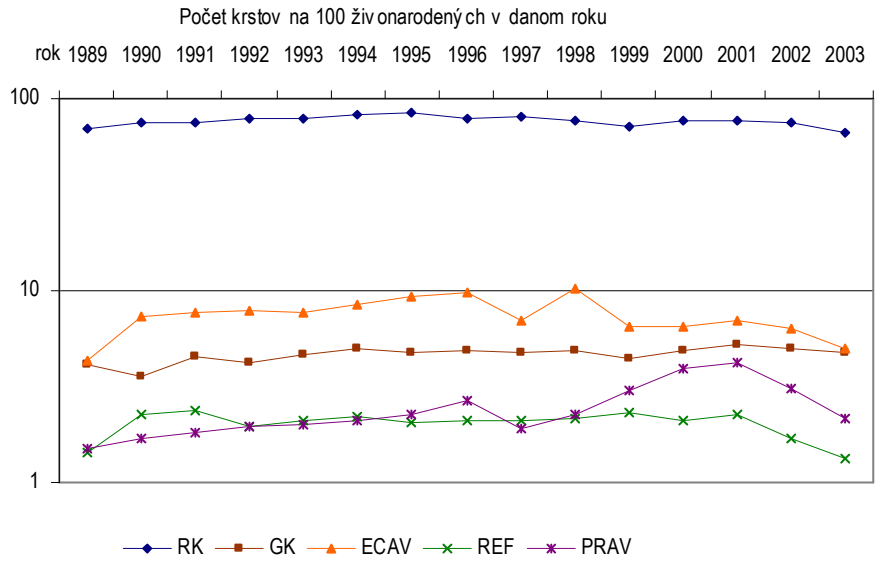
Náboženská príslušnosť obyvateľov Slovenskej republiky bola naposledy zisťovaná pri Sčítaní obyvateľov, bytov a domov v roku 2001. Pod pojmom náboženské vyznanie obyvateľa sa chápala *„účasť na náboženskom živote niektorej cirkvi alebo vzťah k nej“* (1). Vďaka kombinácii údajov zo sčítacích hárkov možno získať prehľad o pohlaví, veku, národnosti, rodinnom stave, atď. príslušníkov jednotlivých vierovyznaní (resp. registrovaných cirkví a náboženských spoločností v danom čase) na Slovensku. Chýbajú nám však dáta, ktoré by zachytili prirodzený pohyb obyvateľov podľa ich konfesií. Túto medzeru možno čiastočne vyplniť údajmi zo Štatistických ročeniek SR (viď *Zdroje dát*), v ktorých sú publikované vybrané ukazovatele pre 5 najpočetnejších registrovaných cirkví na Slovensku: Rímskokatolícku cirkev (RK), Evanjelickú cirkev augsburského vyznania (ECAV), Gréckokatolícku cirkev (GK), Reformovanú kresťanskú cirkev (REF) a Pravoslávnu cirkev (PRAV). Sledovali sme počty krstov a pohrebov, ktoré naznačujú istú líniu prirodzeného pohybu v cirkvi a taktiež počet cirkevných sobášov, pretože proces sobášnosti je u nás aj naďalej výrazným faktorom, ktorý ovplyvňuje reprodukciu obyvateľstva, keďže ešte stále sa väčšina detí rodí vydatým ženám. Použité údaje sú z rokov 1989-2003.

Počet krstov

Údaje o počte krstov nemusia znamenať iba krsty malých detí krátko po narodení väčšinou v danom kalendárnom roku. Po roku 1989 bol počet krstov ovplyvnený i krstmi dospelých. Dokazuje to najmä fakt, že ak spočítame krsty v sledovaných piatich cirkvách, ich súčet je vyšší ako počet živonarodených detí v danom roku (platí to pre rok 1994 a 1995). Sociologička A. Kvasničková zdôvodňuje, že išlo o *„pokrstenie s časovým odstupom od roku narodenia a o konverzie dospelých obyvateľov“* (2). To dokazujú i jej informácie z rímskokatolíckych diecéz.

Údaje o krstoch teda neposkytujú presné informácie o reprodukčnom dianí v piatich najpočetnejších cirkvách (bližšie viď graf A), ale i napriek tomu sú cenným zdrojom informácií o prílive nových členov do cirkvi.

Graf A.



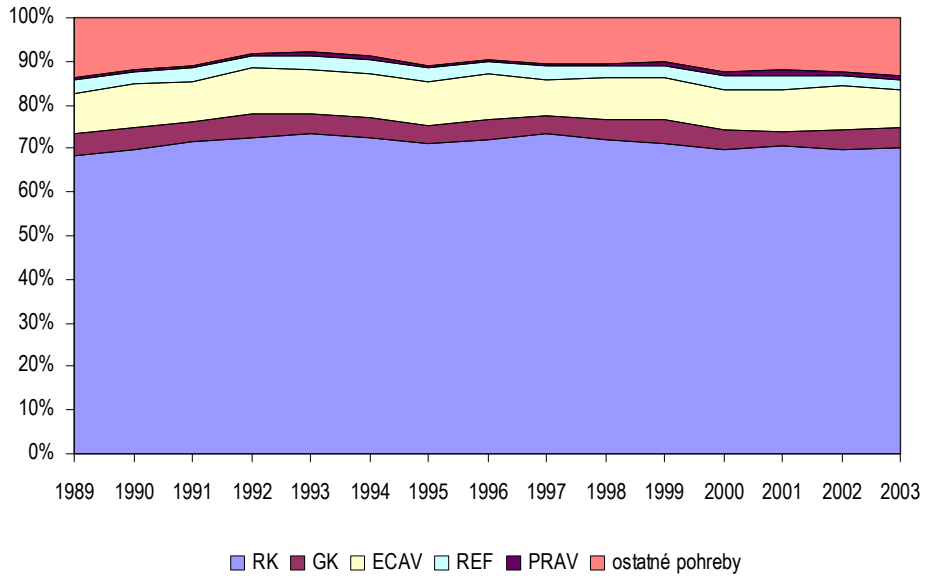
Počet pohrebov

Počet pohrebov taktiež nezachytáva úplne presne úmrtnosť všetkých príslušníkov danej cirkvi, ale je do istej miery presnejší ako sledovanie pôrodnosti prostredníctvom krstov.

Podľa grafu B vidíme, že percentuálny podiel pohrebov v cirkvách sa výrazne nemení a ostáva pomerne stabilný. V rokoch 1989-2003 bolo na Slovensku priemerne 71% pohrebov zo všetkých pohrebov rímskokatolíckych, 10% evanjelických, 5% gréckokatolíckych, 3% kalvínskych, 1% pravoslávnych a zvyšných 10 % tvorili buď necirkevné pohreby alebo pohreby ostatných cirkví. Zaujímavé je, že kým v roku 1996 pripadalo na 100 všetkých pohrebov len 10 necirkevných pohrebov (resp. pohrebov ostatných cirkví), v roku 2003 tento počet vzrástol na 13. V roku 1989 bol síce tento počet 14, ale domnievame sa, že to bolo ovplyvnené nižšou mierou náboženskej slobody.

Graf B.

Pohreby na Slovensku v rokoch 1989-2003



Počet sobášov

Ako sme už v úvode spomínali, proces sobášnosti je stále jeden z faktorov, ktorý vplýva na reprodukciu. „*Vytváraním trvalých alebo dočasných, oficiálnych alebo formálnych partnerských zväzkov vznikajú predpoklady pre proces rodenia detí a v konečnom dôsledku aj pre zmeny početného stavu a štruktúry obyvateľstva*“ (3). K dispozícii v našom prípade sú len údaje o oficiálnych zväzkoch – sobášoch.

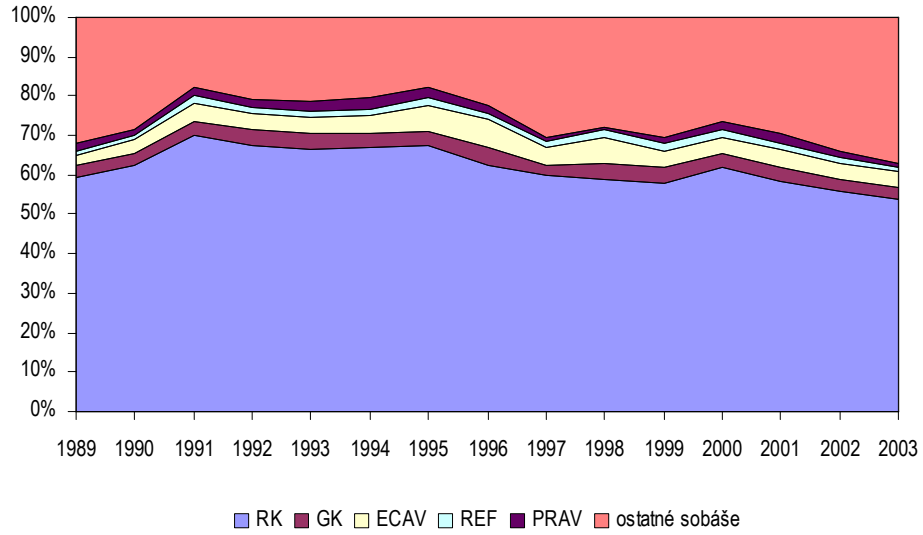
V rokoch 1989-2003 bolo priemerne 62% sobášov rímskokatolíckych, 4% evanjelických, 4% gréckokatolíckych, 2% kalvínskych a 2% pravoslávnych, zvyšných 26% tvorili ostatné sobáše.

Na grafe C vidíme, že podiel ostatných sobášov narastá a svoje maximum dosiahol v roku 2003, keď 37% sobášov nebolo uzavretých v žiadnej z piatich najpočetnejších cirkví na Slovensku. Jednou z príčin, okrem sekularizácie spoločnosti, môže byť aj jednoduchšia a rýchlejšia organizácia sobášov na radniciach, ktorým nepredchádzajú pravidelné prípravné stretnutia na manželstvo s duchovným, ako to je povinnosťou pred uzatvorením cirkevných sobášov.

Do pozornosti treba dať i fakt, že napr. absolútny počet sobášov je vyšší v pravoslávnej cirkvi než v reformovanej kresťanskej cirkvi a to i napriek tomu, že kalvínov je dvakrát viac ako pravoslávnych. Podobne dobieha v sobášnosti aj gréckokatolícka cirkev evanjelickú, pričom gréckokatolícka má o 150 000 príslušníkov menej. Je to pravdepodobne spôsobené najmä odlišnou vekovou štruktúrou a i tzv. kríženými manželstvami, keď sa zosobášia snúbenci s odlišným vierovyznaním. Odlišnú vekovú štruktúru potvrdzuje napr. ukazovateľ vekového mediánu, ktorý bol v roku 2001 najvyšší v evanjelickej cirkvi – takmer 43 rokov, 39 rokov u kalvínov, vo zvyšných troch cirkvách to bolo približne 35 rokov.

Graf C.

Sobáše na Slovensku v rokoch 1989-2003



Záver

Podľa rozdielu medzi sčítaniami obyvateľov v rokoch 1991 a 2001 vieme, že absolútny prírastok svojich vyznávačov zaznamenali všetky nami sledované cirkvi.

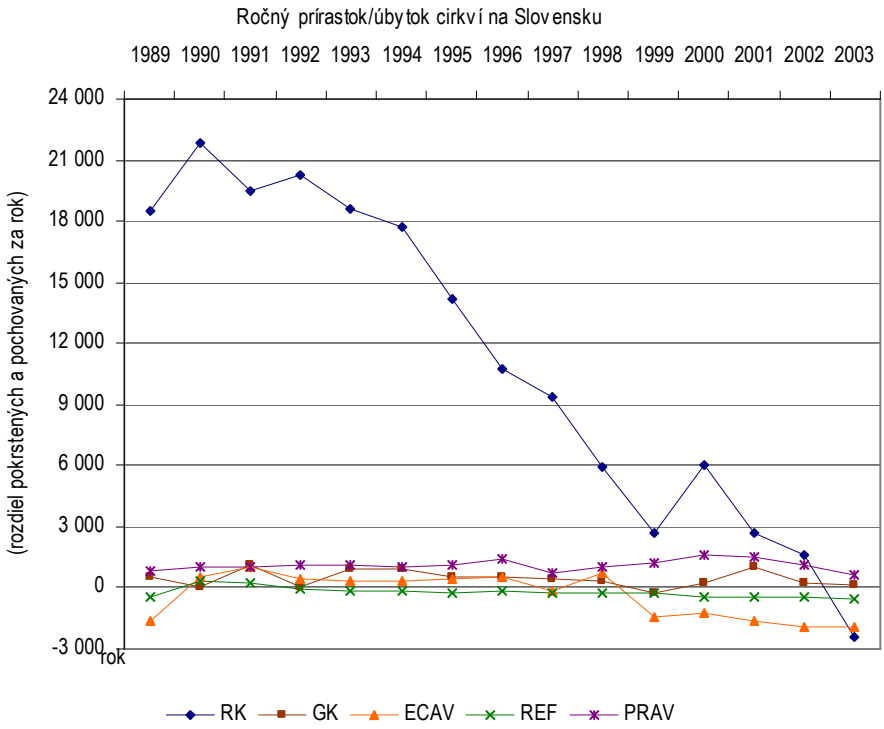
Využili sme naše znalosti o počte krstov a pohrebov v skúmaných cirkvách a snažili sme sa vypočítať teoretický prírastok príslušníkov cirkví na základe rozdielu medzi počtom pokrstených a pochovaných – snaha o zachytenie prirodzeného pohybu. K základnému údaju o počte veriacich vybranej cirkvi z 3.3.1991 sme pripočítali 10/12 prírastku, resp. úbytku (rozdiel počtu krstov a pohrebov) v roku 1991, ďalej sme pripočítali celé prírastky, resp. úbytky za roky 1992-2000 a napokon sme pripočítali 5/12 prírastku, resp. úbytku za rok 2001 a porovnali sme ho s výsledkom zo sčítania z 26.5.2001. Výsledné čísla tohto pokusu o zachytenie prirodzeného prírastku na základe počtu krstov a pohrebov sú výrazne nižšie ako výsledky zo sčítania v roku 2001. Medzicenzový príliv príslušníkov piatich cirkví teda nebol spôsobený vysokou pôrodnosťou, resp. nízkou úmrtnosťou členov cirkví, ale išlo o obyvateľov, ktorí uviedli svoje vierovyznanie až v roku 2001 a prešli z kategórie nezistených v roku 1991 do kategórie príslušníkov nejakej cirkvi v roku 2001. K danej cirkvi sa mohli prihlásiť jednak pokrstení pred rokom 1991, ale i sympatizanti s cirkvou (nemusia byť pokrstení), ktorí sa častokrát príliš nezúčastňujú náboženských obradov, resp. úkonov a s cirkvou ich do istej miery spája napr. rodinná tradícia a pod.

Počet príslušníkov cirkví	*Cenzus v roku 1991*	*Cenzus v roku 2001*	*Teoretický počet v roku 2001 podľa ročných prírastkov/úbytkov*
Rímskokatolícka cirkev	3 187 383	3 708 120	3 310 247
Evanjelická a. v. cirkev	326 397	372 858	326 463
Gréckokatolícka cirkev	178 733	219 831	183 802
Reformovaná kresťanská cirkev	82 545	109 735	80 691
Pravoslávna cirkev	34 376	50 363	46 194

Graf D ďalej ozrejmuje, že ročný prírastok členov cirkví (rozdiel počtu krstov a počtu pohrebov) v časovom období 1989-2003 celkovo klesá, ba u reformovanej, evanjelickej a aj rímskokatolíckej cirkvi bol počet pohrebov vyšší ako prírastok príslušníkov zaznamenaný prostredníctvom krstov. Tento trend naznačuje, že počet obyvateľov hlásiacich sa k 5 najpočetnejším

cirkvám v ďalšom sčítaní bude približné rovnaký, resp. možno očakávať poklesov počtu príslušníkov skúmaných cirkví.

Graf D.



Zdroje dát:
Štatistická ročenka Slovenskej republiky 2004. Štatistický úrad Slovenskej republiky, VEDA – vydavateľstvo SAV, Bratislava 2004 (str. 199, 551-552)
Štatistická ročenka Slovenskej republiky 1999. Štatistický úrad Slovenskej republiky, VEDA – vydavateľstvo SAV, Bratislava 1999 (str. 171, 518-519)
Štatistická ročenka Slovenskej republiky 1994. Štatistický úrad Slovenskej republiky, VEDA – vydavateľstvo SAV, Bratislava 1995 (str. 130, 420-421)

Použitá literatúra:

(1) http://www.statistics.sk/webdata/slov/scitanie/opatsusr.htm
(2) Kvasničková, A. Náboženský život. In: Sopóci, J. a kol.: Slovensko v deväťdesiatych rokoch. Vydavateľstvo UK Bratislava 2003, (str. 267-308).
(3) Kol. autorov: Populačný vývoj v Slovenskej republike 2002. Akty Bratislava 2003 http://www.infostat.sk/vdc/pdf/popul2002x.pdf

Viera Hluchá
Univerzita Komenského Prírodovedecká fakulta,
5. ročník, odbor geografia - kartografia

Metódy viackriteriálneho hodnotenia a ich aplikácia na finančné ukazovatele kapitálových spoločností

Viera Justiňáková

Cieľ článku

Hlavnou úlohou tohto článku je priniesť stručný prehľad metód, ktoré sa môžu využiť pri porovnávaní finančnej situácie kapitálových spoločností. Ďalším cieľom je porovnať výsledky jednotlivých metód a na ich základe sa pokúsiť vytvoriť rebríček úspešnosti spoločností s ručením obmedzeným a akciových spoločností.

Úvod

Každý podnik je špecifickou, jedinečnou jednotkou, ktorú nie je jednoduché porovnávať s iným podnikom, i v prípade keď odvetvie ich podnikania je rovnaké. Výsledky hospodárenia každého podniku sú ovplyvňované prostredím, v ktorom vykonáva svoju podnikateľskú činnosť. V rámci zrýchľujúceho sa procesu globalizácie a zosilňujúceho tlaku konkurencie, je dôležité poznať svoju pozíciu v podnikateľskom prostredí. Práve metódy viackriteriálneho hodnotenia nám umožňujú porovnávať akýkoľvek objekt na základe niekoľkých ukazovateľov.

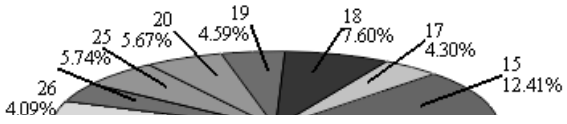
Popis údajov

Metódy viackriteriálneho hodnotenia budeme aplikovať na údajovú základňu, ktorej prvú časť tvoria ekonomické ukazovatele charakterizujúce hospodárenie spoločností s ručením obmedzeným za rok 2003. Druhú časť tvoria ekonomické ukazovatele, ktoré charakterizujú hospodárenie akciových spoločností za rok 2003. Všetky údaje sa týkajú priemyselných spoločností, sídliacich na území Slovenskej republiky s počtom pracovníkov vyšším ako 20. Údaje pochádzajú zo Štatistického úradu Slovenskej republiky a sú určené pre vedecké účely. Názvy podnikov sú anonymné, pričom jednotlivé spoločnosti sú bližšie identifikované okresom, v ktorom sa nachádzajú, presnejšie číslom okresu podľa štatistických číselníkov okresov a druhom činnosti, ktorú vykonávajú a to podľa odvetvovej klasifikácie ekonomickým činností.

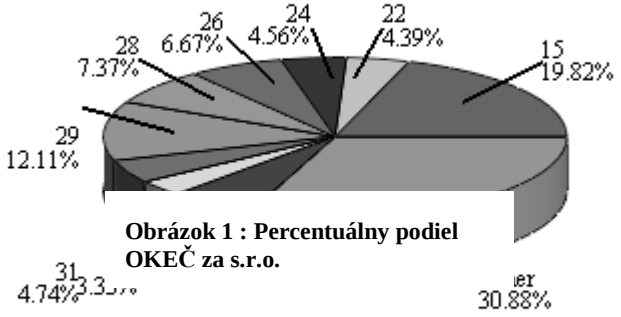
Na lepšie pochopenie štruktúry údajov využijeme grafickú analýzu, konkrétne graf „Pie Chart“, známy pod názvom koláčový graf. Našou úlohou bude znázorniť percentuálne zastúpenie jednotlivých druhov ekonomických činnosti, ktoré sú očíslované podľa číselníka OKEČ.

Najväčšiu časť na obrázku 1 a obrázku 2 zaberá položka s názvom „Other”. Ide o položku , v ktorej sú zahrnuté podniky vykonávajúce takú ekonomickú činnosť, ktorej zastúpenie neprekročilo hodnotu 3%. V rámci údajov o s.r.o. sú percentuálne podiely jednotlivých ekonomických činností veľmi tesné. Na obrázku 1 vidíme, že najväčšie zastúpenie až 14,78% majú podniky, ktoré sa venujú výrobe kovových konštrukcií a kovových výrobkov okrem výroby strojov a zariadení. Podľa číselníka Odvetvovej klasifikácii ekonomických činností, je táto činnosť označená číslom 28. V rámci údajov o a.s., najväčší percentuálny podiel majú spoločnosti, ktoré sa venujú výrobe potravín a nápojov (OKEČ15) s podielom 19,82%.

s.r.o.



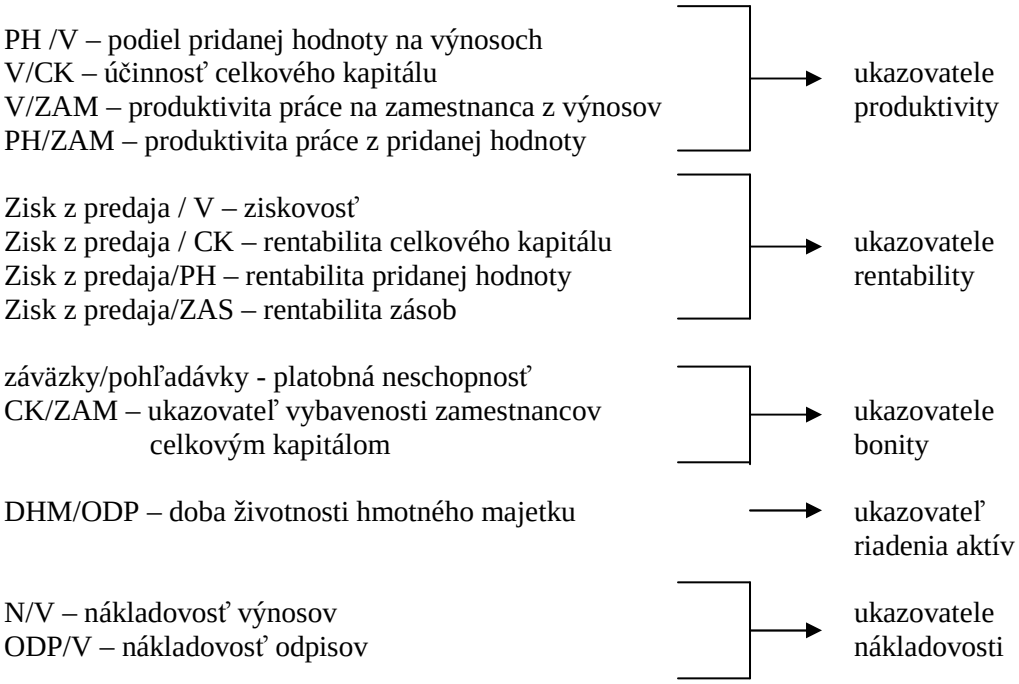
a.s.



Obrázok 1 : Percentuálny podiel OKEČ za s.r.o.

Hospodárenie obidvoch typov spoločností je v danej údajovej základni charakterizované jedenástimi ekonomickými ukazovateľmi: výnosmi (V), nákladmi(N), pridanou hodnotou(PH), odpismi(ODP), ziskom z predaja, pohľadávkami, záväzkami,

celkovým kapitálom(CK), zamestnancami(ZAM), dlhodobým hmotným majetkom(DHM) a zásobami(ZAS). Všetky ekonomické ukazovatele vyjadrujú stav na konci roku 2003. V prvom kroku spracovania údajov, si pôvodné ekonomické ukazovatele zosumarizujeme za jednotlivé okresy, čím zmenšíme rozsah súboru a spravíme našu analýzu zaujímavejšou. Potom z týchto zosumarizovaných ukazovateľov vytvoríme relatívne finančné ukazovatele, nakoľko pre porovnanie je vhodnejšie pracovať s relatívnymi ukazovateľmi. Ďalej budeme hodnotiť a porovnávať jednotlivé okresy, na základe relatívnych finančných ukazovateľov. To znamená, že za jednotlivé okresy vytvoríme relatívne finančné ukazovatele, na ktoré sa aplikujú metódy viackriteriálneho hodnotenia. Pri spoločnosti s ručením obmedzeným aj pri akciovej spoločnosti som vytvorila 13 finančných ekonomický ukazovateľov.



Metódy viackriteriálneho hodnotenia

Cieľom metód viackriteriálneho hodnotenia je transformácia a syntetizácia hodnôt všetkých ukazovateľov do jednej výslednej charakteristiky, ktorá vyjadruje komplexne úroveň jednotlivých objektov. Medzi metódy viackriteriálneho hodnotenia zaraďujeme:

A. Metódu váženého súčtu poradí
B. Bodovaciu metódu
C. Metódu normovanej premennej
D. Metódu vzdialenosti od fiktívneho objektu

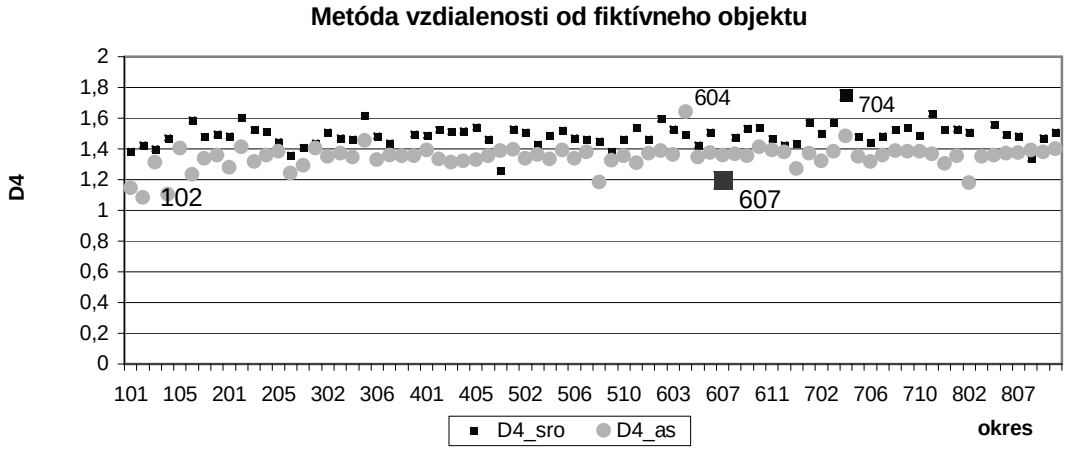
Metódy viackriteriálneho hodnotenia sa navzájom odlišujú v čiastkových hodnoteniach, čiže podľa jednotlivých ukazovateľov a spôsobom ich zhrnutia do celkového hodnotenia. Voľba metódy viacrozmerného hodnotenia závisí od rozhodnutia samotného analytika.

Pri analýze je vhodné použiť viaceré metódy a na záver porovnávať ich výsledky. Tento postup som použila aj ja pri svojej analýze. Po usporiadaní pôvodného štatistického súboru podľa okresov, som analyzovala 77 okresov pri údajoch o s.r.o. a 76 okresov pri údajoch o a.s..[1] Po aplikácii metódy vzdialenosti od fiktívneho objektu na relatívne finančné ukazovatele , som integrálnemu ukazovateľu D4 za každý okres, priradila príslušné poradie. Rovnako som postupovala aj pri ostatných troch metódach, avšak na interpretáciu som vybrala výsledky metódy vzdialenosti od fiktívneho objektu, ktorá sa odlišuje od ostatných tým, že najlepšie výsledky bude dosahovať ten objekt (okres), ktorého

[1] Údajová základňa neobsahuje údaje o priemyselných podnikoch zo všetkých okresov Slovenska, pričom údaje o s.r.o. obsahujú o jeden okres viac oproti a.s., preto posledné miesto rebríčka úspešnosti pri s.r.o. bude mať číslo 77 a pri a.s. číslo 76.

integrálny ukazovateľ bude dosahovať najnižšiu hodnotu. Zároveň táto metóda pracuje so štvorcami odchýlok, čiže výsledky nadobúdajú vždy kladné hodnoty.

Na obrázku 3 sú znázornené hodnoty integrálneho ukazovateľa D4 . Na horizontálnej osi grafu sú znázornené čísla jednotlivých okresov a na osi vertikálnej sú hodnoty ukazovateľa D4. Čierne štvorce predstavujú hodnoty integrálneho ukazovateľa D4 spoločností s ručením obmedzeným a šedé kruhy predstavujú hodnoty D4 akciových spoločností.



Obrázok 3 : Hodnoty integrálneho ukazovateľa D4

Z obrázku 3 vieme určiť poradie jednotlivých spoločností. Medzi podnikmi s ručením obmedzeným prvé miesto, čiže najlepšie výsledky podľa metódy vzdialenosti od fiktívneho objektu má okres č.607 - Okres Poltár. Hodnota jeho integrálneho ukazovateľa D4 bola v porovnaní s ostatnými okresmi najnižšia. Posledné miesto obsadil okres č. 704 – Okres Levoča. Hodnota jeho integrálneho ukazovateľa D4 bola v porovnaní s ostatnými okresmi najvyššia.

Medzi akciovými spoločnosťami sa na prvom mieste umiestnil okres Bratislava II s číslom 102 a na poslednej pozície je okres Detva s číslom 604. Všetky tieto okresy sú na obrázku 3 vyznačené prislúchajúcim číslom okresu.

Vyhodnotenie

Výsledné tabuľky obsahujú prvé, druhé, tretie a posledné miesto, na ktorom sa umiestnili tie okresy, z ktorých pochádzajú spoločnosti dosahujúce, podľa metód viackriteriálneho hodnotenia, najlepšie resp. najhoršie výsledky, v porovnaní s ostatnými spoločnosťami. Podrobný popis hodnotenia všetkých okresov vzhľadom k rozsiahlosti výslednej tabuľky neuvádzam.

Metóda váženého súčtu poradí		
	s.r.o.	a.s.
Poradie	okres	
1.	Košice II	Bratislava I
2.	Skalica	Košice I
3.	Bratislava I	Dunajská Streda
77. (76.)	Kežmarok	Levoča

Bodovacia metóda		
	s.r.o.	a.s.
Poradie	okres	
1.	Poltár	Bratislava I
2.	Košice II	Bratislava II
3.	Zlaté Moravce	Bratislava IV
77. (76.)	Levoča	Detva

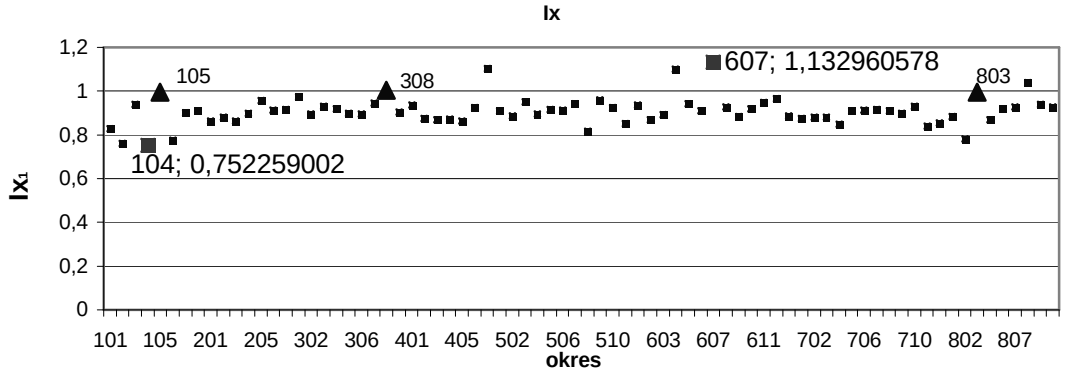
Metóda normovanej premennej		
	s.r.o.	a.s.
Poradie	okres	
1.	Poltár	Bratislava IV
2.	Košice II	Bratislava I
3.	Zlaté Moravce	Bratislava II
77. (76.)	Levoča	Detva

Metóda vzd. od fiktívneho objektu		
	s.r.o.	a.s.
Poradie	okres	
1.	Poltár	Bratislava II
2.	Zlaté Moravce	Bratislava IV
3.	Rožňava	Bratislava I
77. (76.)	Levoča	Detva

Z výsledných tabuliek môžeme urobiť nasledovný záver. Spoločnosť s ručením obmedzeným, ktorá dosiahla v roku 2003 najlepšie výsledky, pochádza z okresu Poltár a venuje sa výrobe kovov. Takto konkrétne vieme definovať daný podnik z toho dôvodu, že naša údajová základňa zaznamenáva iba jednu spoločnosť s ručením obmedzeným v okrese Poltár.
O najúspešnejších akciových spoločnostiach vieme povedať, že sa nachádzajú v okrese Bratislava I, II alebo IV, nakoľko sa tieto okresy v troch metódach viackriteriálneho hodnotenia striedajú na prvých troch priečkach rebríčka úspešnosti. Odchýlky medzi prvými troma pozíciami sú v týchto metódach také malé, že nepovažujem predchádzajúce výpočty za nepresné alebo chybné, aj keď ich výsledky nie sú totožné. Výsledky metódy váženého súčtu poradí sa odlišujú od ostatných troch metód aj pri spoločnostiach s ručením obmedzeným aj pri akciových spoločnostiach. Tento rozdiel pripisujem až prílišnej jednoduchosti danej metódy a veľkosti súboru, pričom prednosť dávam výsledkom získaným na základe ostatných troch metód.

Porovnanie výsledkov

Zaujímavé je zistiť, pomocou relatívneho porovnania výsledkov metódy vzdialenosti od fiktívneho objektu, ako sa od seba odlišujú akciové spoločnosti a spoločnosti s ručením obmedzeným v jednotlivých okresoch. Táto problematika je vyznačená na obrázku 4.



Obrázok 4: Porovnanie D4 medzi okresmi

Na horizontálnej osi sú vyznačené jednotlivé okresy a na vertikálnej osi je hodnota indexu Ix, ktorú som vypočítala podľa vzťahu (1).

$$I_x = \frac{x_k}{x_j} \qquad (1)$$

x_k - hodnota D4 akciových spoločností, usporiadaná podľa okresov $k = 1,2,\ldots,76$
x_j - hodnota D4 spoločností s ručením obmedzeným, usporiadaná podľa okresov $j = 1,2,...,76$

Z grafu vieme vyčítať, že v okrese Poltár dosahujú spoločnosti s ručením obmedzeným o 13,3% lepšie výsledky ako akciové spoločnosti. V okrese Bratislava IV majú akciové spoločnosti, podľa výsledkov metódy vzdialenosti od fiktívneho objektu, o 24,77% lepšie výsledky oproti spoločnostiam s ručením obmedzeným. Čísla týchto okresov sú zobrazené na obrázku 4 aj s hodnotou prislúchajúceho indexu, pričom sa jedná o okresy, ktoré dosiahli podľa metódy vzdialenosti od fiktívneho objektu najlepšie výsledky. Z okresu Poltár pochádza najúspešnejšia spoločnosť s ručením obmedzeným a z okresu Bratislava IV najlepšie akciové spoločnosti. Z grafu vidíme, že existujú aj také okresy, kde vychádzajú hodnoty integrálneho ukazovateľa skoro rovnaké, čiže hodnota indexu I_x je približne rovná jednej. Tieto hodnoty sú na obrázku 4 znázornené trojuholníkom a prislúchajúcim číslom okresu. Konkrétne ide o okres Bratislava V (č.105), okres Púchov (č.308) a okres Košice II (č.803).

Záver

V súčastnej rýchlej dobe je veľmi dôležité dokázať poznať svoje prostredie, ohodnotiť vlastné postavenie a porovnať sa s okolím. Na porovnanie rôznych druhov objektov, na základe niekoľkých ukazovateľov nám slúžia metódy viackriteriálneho hodnotenia. Spoľahlivosť jednotlivých metód sme si ukázali na konkrétnom príklade. Metóda váženého súčtu poradí je síce jednoduchá, ale pri veľkom počte objektov nie príliš presná, pričom výsledky ostatných troch metód sa približne zhodujú. Na základe výsledkov týchto metód a pomocou indexov vieme urobiť rôzne druhy porovnaní, pomocou ktorých si dokáže podnik určiť svoju pozíciu na trhu.

Zoznam použitej literatúry:

1. Zalai, K.: Finančno-ekonomická analýza. Bratislava, Sprint vfra 2000.

2. Chajdiak, J.: Ekonomická analýza stavu a vývoja firmy. Bratislava, Statis 2004.

3. Pacáková, V.: Štatistika pre ekonómov. Iura edition, 2003 .

4. Šnircová, J.: Porovnávate finančné výsledky svojho podniku s konkurenciou v odvetví? Finančný manažér č. 2/2005.

Viera Justiňáková
Fakulta Hospodárskej Informatiky EU Bratislava, 5. ročník
Kontakt na autora : e-mail : viera@cis.sk

Zmeny a nové formy rodinného správania na Slovensku
Kohabitácie ako jedna z foriem partnerského spolužitia

TATIANA LACHOVÁ

Demografický vývoj a rodina sú nepochybne vo veľmi úzkom vzťahu. Rodina bez ohľadu na civilizačné, kultúrne a v tom obsiahnuté i hodnotové zmeny, je stále základnou reprodukčnou jednotkou spoločnosti. Rodina na Slovensku sa na základe poznatkov získaných z projektu EVS (European values system, v prekl. „Dotazník európskych hodnôt“) uskutočneného v rokoch 1991 a 1999 a prieskumu populačných zámerov vysokoškolských študentov v kontexte druhej demografickej revolúcie realizovaného Katedrou sociológie FF UK a Ústavu sociálneho lekárstva LF UK v roku 1997 ukazuje byť stabilnou hodnotou v živote ľudí, napriek tomu sa však mení. Potvrdzujú to zmeny sledované prostredníctvom indikátorov ako sú napr. založenie rodiny, generačné bývanie, zdroj obživy, počet detí, starostlivosť o starých členov rodiny a pod. Tieto zmeny sú dôsledkom meniacich sa podmienok, ktoré tvoria vonkajšie prostredie rodiny a patria sem hlavne meniace sa ekonomické, politické, demografické a societálne štruktúry spoločnosti.

V posledných rokoch, okrem iného v období poklesu sobášnosti, sa v spoločnosti rozšíril nový fenomén - kohabitácie. Pri SĽDB 1991 sa zisťoval ich počet v populácii. Celkovo bolo v roku 1991 na Slovensku 20 864 kohabitácií. Tento počet však nebol úplný. Na počiatku 90. rokov bolo prehlásenie vzťahu druh-družka oveľa chúlostivejšou záležitosťou, než je tomu v súčasnej dobe, naviac sa zápis o nezosobášenom súžití týkal výhradne párov, kde sa obidvaja partneri mohli preukázať svojím trvalým bydliskom. Väčšinou preto šlo o zväzky strednej a staršej generácie, ktoré boli alternatívou manželstva z obavy nad stratou vdovského dôchodku. Do roku 2001 sa počet kohabitácií výrazne zvýšil na 30 466. Pri všetkých štatistických cenzoch je však otázne, koľko kohabitantov svoje spolužitie prizná.

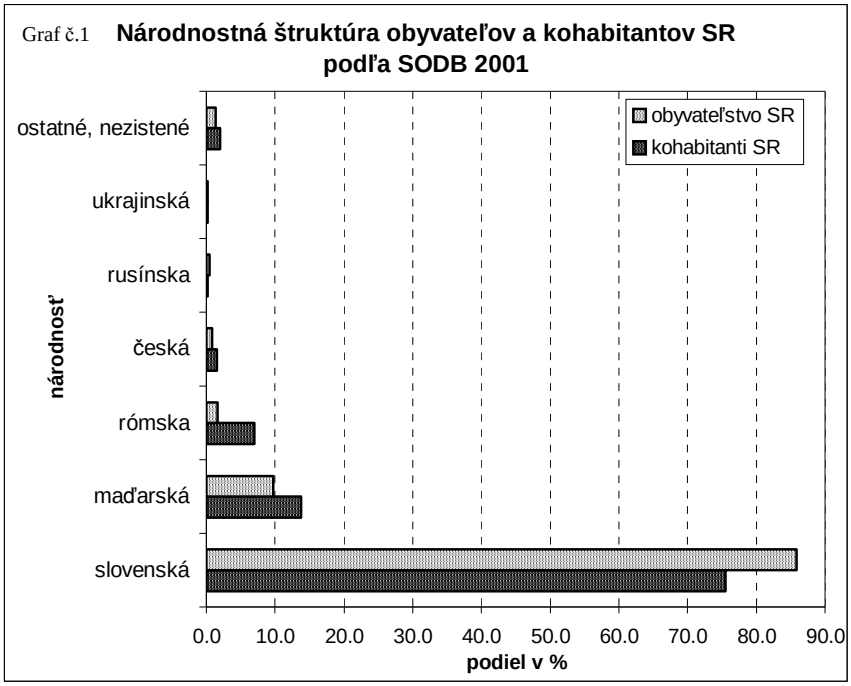
Tento fakt je už výsledkom zmien rodinného správania obyvateľstva; spoločnosť začína rešpektovať odlišné formy partnerského života. Ich rozšírenie v globálnom meradle je odrazom komplexných zmien spoločenských a ekonomických podmienok, zahŕňajúcich zmeny vzorov zamestnanosti a obsahu práce, zmeny v oblasti vzdelávania, konkrétne rast vzdelanostnej úrovne žien, meniace sa vzťahy medzi mužmi a ženami, posuny v hodnotových preferenciách a očakávaniach, globalizačné procesy, pokles tradičných inštitúcií poskytujúcich sociálne zabezpečenie, snaha žien budovať si finančnú nezávislosť, rast pracovných a kariérových ašpirácií, zlepšenie celkových pozícií žien predovšetkým na trhu práce, nové možnosti, ktoré prispeli k väčšej osobnej autonómii, sebaidentite a sebanaplneniu prostredníctvom zapojenia sa do platenej práce, ktorá bola predtým doménou mužov.

Vplyv sociokultúrnych faktorov na formovanie kohabitácií

Z hľadiska charakteristických znakov rodiny, niektoré sociokultúrne faktory daného spoločenstva či niektorých sociálnych skupín, môžu dostávať regionálnu dimenziu. Medzi takéto faktory, majúce i regionálnu dimenziu, do istej miery patrí napr. národnosť, religiozita a vzdelanie. Tieto sa premietajú do diferencii v rodinách, opierajú sa o odlišné tradície pri zakladaní rodiny, majú vplyv na výchovu detí i veľkosť rodiny. Národnostná štruktúra, religiózna príslušnosť ako i vzdelanostná úroveň obyvateľstva patria práve medzi tie sociokultúrne faktory, ktoré sú v pomerne zložitých vzťahoch a väzbách s rodinným správaním obyvateľstva a zároveň i s úrovňou kohabitácií.

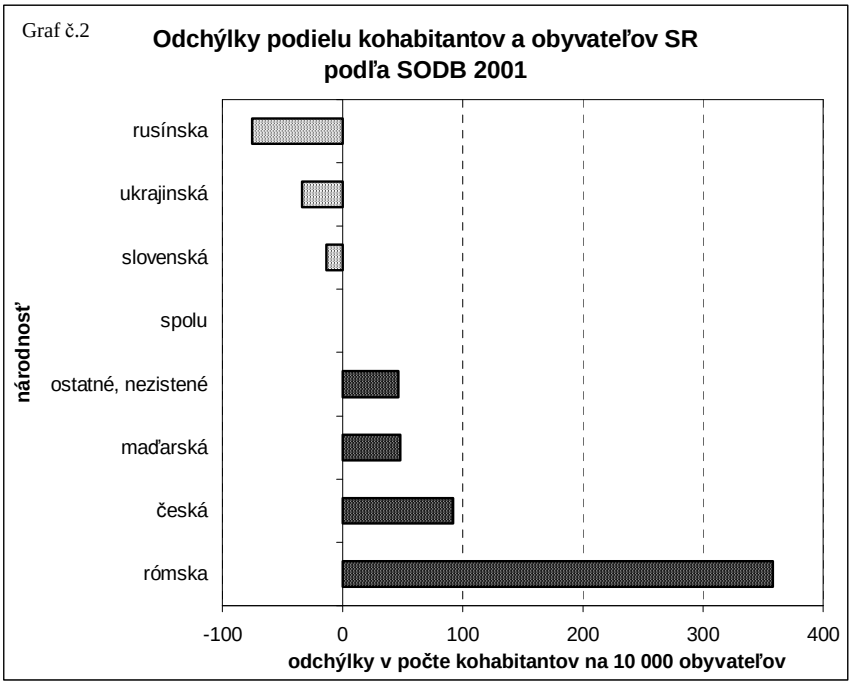
Národnostná štruktúra kohabitantov v SR podľa SODB 2001

Graf č.1 **Národnostná štruktúra obyvateľov a kohabitantov SR podľa SODB 2001**



Z kultúrnych charakteristík sa národnosť obyvateľstva prejavuje ako určitý syntetický znak s výraznejším vplyvom na populačné javy a procesy (Mládek, J., 2003). Môžeme to sledovať i pri hodnotení kohabitácií a národnostnej štruktúry kohabitantov v Slovenskej republike. Výsledky komparácie národnostnej štruktúry kohabitantov a obyvateľov SR poukazujú na relatívnu podobnosť oboch súborov. Už i menšie odlišnosti týchto štruktúr však treba považovať za významné, pretože práve tieto rozdiely umožňujú vysvetliť vzťah príslušníkov jednotlivých národností ku kohabitáciám. Pri národnostnej štruktúre kohabitantov ako i obyvateľov SR, celkovo najpočetnejšie zastúpenie dosahujú osoby slovenskej, maďarskej a rómskej národnosti (viď graf č.1).

Graf č.2 **Odchýlky podielu kohabitantov a obyvateľov SR podľa SODB 2001**



V snahe exaktnejšie vyjadriť podiely kohabitantov jednotlivých národností, boli dané do pomeru počty kohabitantov a obyvateľov jednotlivých národností Slovenska a stanovil sa tak ukazovateľ počtu kohabitantov na 10 000 obyvateľov (priemer za Slovensko predstavuje 113,3). Výsledkom sú odchýlky týchto ukazovateľov vyjadrené pre jednotlivé národnosti. Relatívne nižšiu intenzitu kohabitácií, teda určitú zdržanlivosť ba až odmietavý postoj voči kohabitáciám, možno pozorovať u troch národností - slovenskej,

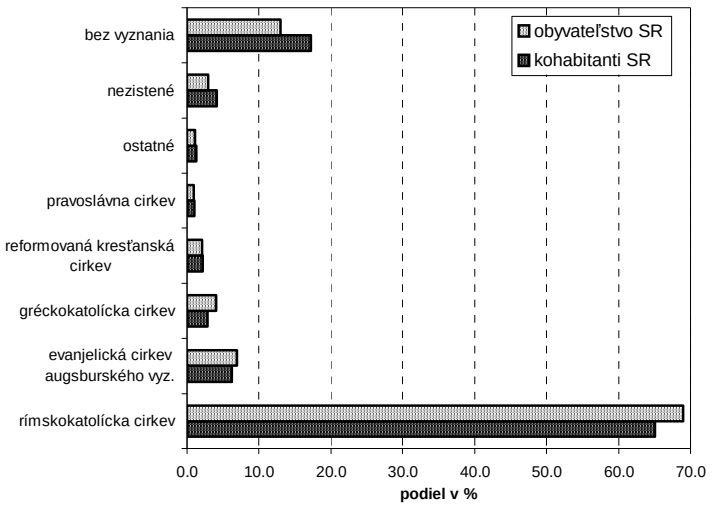
ukrajinskej a rusínskej. Zvýšenú intenzitu kohabitácií a teda pozitívne odchýlky zaznamenávajú príslušníci rómskej, českej a maďarskej národnosti a ostatných, nezistených národností (viď graf č.2).

Religiózna štruktúra kohabitantov v SR podľa SODB 2001

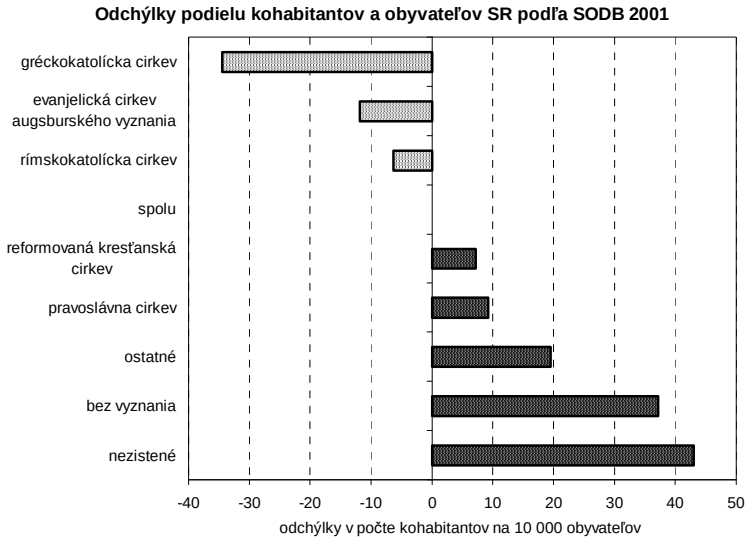
Napriek postupujúcej sekularizácii sa vplyv religióznej príslušnosti pociťuje veľmi výrazne vo viacerých demografických procesoch. Podobne aj rodinné správanie odráža postoje jednotlivých relígií k procesom formovania, existencie a rozpadu rodiny (Mládek, J., 2003).

Podľa religióznej štruktúry kohabitantov v SR v roku 2001 najväčšie zastúpenie v kohabitáciách majú rímskokatolíci, osoby bez vyznania, evanjelici augsburského vyznania, gréckokatolíci a reformovaní kresťania (viď. graf č.3). V štruktúre kohabitantov Slovenska sa s pozitívnymi odchýlkami prejavujú osoby s nezisteným vierovyznaním, bez vyznania a patriaci k ostatným relígiám. S malými pozitívnymi odchýlkami sa prezentujú i kohabitanti prislúchajúci k pravoslávnej a reformovanej kresťanskej cirkvi. Na rozdiel od toho, určitý zdržanlivý až odmietavý postoj ku kohabitáciám vykazujú stúpenci troch najsilnejšie zastúpených relígií na Slovensku - rímskokatolíckej, gréckokatolíckej a evanjelickej cirkvi augsburského vyznania, pričom najmä váha príslušníkov rímskokatolíckej cirkvi je dominantná. Kohabitanti týchto vyznaní majú relatívne menšiu početnosť a vykazujú tak negatívne odchýlky (viď graf č.4).

Graf č.3 **Religiózna štruktúra obyvateľov a kohabitantov SR podľa SODB 2001**



Graf č.4



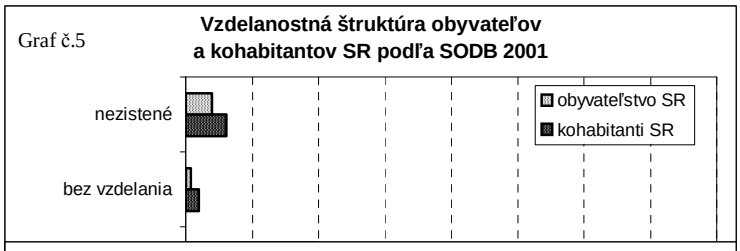
Vzdelanostná úroveň kohabitantov v SR podľa SODB 2001

Hodnotenie vzťahu medzi úrovňou vzdelania a intenzitou kohabitácií je dosť komplikované a vývojovo sa mení. Zahraničné pramene potvrdzujú priamy vzťah medzi úrovňou vzdelania a intenzitou kohabitácií, ktorý platil v 60. a 70. rokoch 20. storočia. Najčastejšie sa tento znak spomína s úrovňou vzdelania žien. Ženy s nižším vzdelaním sa oveľa ťažšie rozhodujú pre neformálne spolužitie než ženy s vysokoškolským vzdelaním. Preto je účasť vzdelaných žien na kohabitáciách početnejšia, pričom sa uplatňuje najmä určitá výhoda vo forme ich ekonomickej nezávislosti (Mládek, J., 2003).

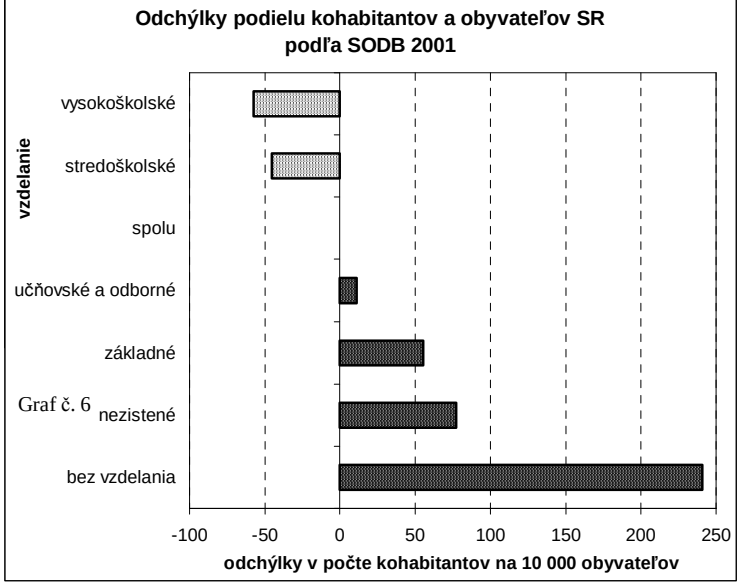
V ďalšom vývoji, počas 80. a 90. rokoch minulého storočia, sa vzťah vzdelania obyvateľstva a kohabitácie mení najmä v dôsledku zmien vekovej štruktúry kohabitantov. Znižuje sa ich vek a do kohabitácií vstupujú stále mladšie vekové kategórie. Je preto logické, že majú ukončené iba nižšie stupne vzdelania alebo sa ešte stále pripravujú na svoje povolanie.

Pri hodnotení vzdelanostnej úrovne obyvateľov a kohabitantov Slovenska sa potvrdili podobné vývojové trendy. Najviac kohabitácií tvoria osoby s najvyšším dosiahnutým stupňom vzdelania základným a učňovským či odborným (viď graf č.5). Podieľajú sa 68,3 % na celkovom počte kohabitácií, zatiaľ čo podiel obyvateľov Slovenska s rovnakým vzdelaním je iba 55,8 %. Odrážajú to i porovnávacie charakteristiky s pozitívnymi odchýlkami (viď graf č.6). Príčinou je narastanie početnosti kohabitácií v nižších vekových kategóriách i početnejšie zastúpenie obyvateľov rómskej národnosti, ktorého vzdelanostná úroveň je nízka. Logicky najvyššie hodnoty pozitívnej odchýlky dosahujú kohabitanti

Graf č.5



Graf č. 6



bez vzdelania alebo s nezisteným vzdelaním. Kohabitanti so stredoškolským a vysokoškolským vzdelaním nedosahujú proporcionálne zastúpenie a charakteristické sú preň záporné odchýlky.

Špecifické črty chápania partnerského spolužitia u Rómov

Kohabitácia predstavuje významný podiel aj na mimomanželskej plodnosti zistenej u rómskeho etnika. Tento údaj je však potrebné doplniť vysvetlením kultúrneho kontextu. Výrazne pozitívna odchýlka v intenzite kohabitácií obyvateľov rómskej národnosti súvisí s osobitosťami zakladania a existencie manželstva a rodiny tohto obyvateľstva. Vznik matrimónia (manželstva) je v ich spoločenstve podriadený vnútroskupinovým zvyklostiam a normám. Rómovia na Slovensku ešte na mnohých miestach zachovávajú, prípadne oživujú tradičnú prax svadobného obradu *mangavipen* (doslova „pytačky, zásnuby"), ktorý sa uskutočňuje vo veľmi mladom veku partnerov a z celospoločenského hľadiska nie je inštitucionálnou legitimizáciou vzťahu muža a ženy; prípadne realizujú vznik manželstva iným, avšak v komunite akceptovaným spôsobom – útekom k príbuzným do inej dediny. V tomto prípade sa *mangavipen* môže, ale nemusí uskutočniť. Dvojicu zosobášenú tradičným spôsobom rómska komunita akceptuje ako manželov a ich deti ako legitímne. Civilný alebo cirkevný sobáš (*civilno vera* resp. *solacharel*, *andre khangeri vera*) nasleduje často až niekoľko rokov po obrade *mangavipen*, zväčša sa viaže na želanie rodičov pokrstiť deti. Po cirkevnom sobáši nasleduje samotné svadobné veselie - *bijav*. Na rozdiel od obradu *mangavipen*, táto slávnosť obsahuje množstvo obyčajových úkonov prevzatých od majoritného obyvateľstva. Práve časový rozdiel medzi vznikom partnerského spolužitia v zmysle vnútroetnických noriem a formálnym uzavretím manželstva vysvetľuje prítomnosť relatívne početnejšieho súboru kohabitácií u rómskeho obyvateľstva.

Ďalšou odlišujúcou charakteristikou tohto etnika je, že v rómskej tradícii je muž či žena dospelý v okamihu, keď je fyzicky schopný založiť si rodinu. Preto je veľmi často prestupovaná norma zákazu sexuálneho styku s osobou mladšou ako 15 rokov, a tak rómske ženy začínajú rodiť vo veľmi nízkom veku a matkami sa stávajú často pred dosiahnutím dospelosti. Nízky vek matiek pri narodení dieťaťa je jednou z príčin, ktorá im bráni uzatvárať manželstvo a s otcom dieťaťa matka žije v nezosobášenom vzťahu.

Literatúra:

[1] BAČOVÁ, V. (1990). Typológia rómskych rodín na Slovensku. Veda, Bratislava
[2] BENEŠOVÁ, V. (2001). Současné demografické změny podle výsledků sociologických výzkumů. Demografie, roč. 43, č. 2
[3] HAVIAROVÁ, E. (2001). Demografická charakteristika Rómskeho obyvateľstva, prípadová štúdia regiónu Horehronie. Diplomová práca. Katedra humánnej geografie a demogeografie, Prírodovedecká fakulta UK v Bratislave, Bratislava
[4] HORVÁTHOVÁ, E. (1964). Cigáni na Slovensku. Historicko – etnografický náčrt. Vydavateľstvo Slovenskej akadémie vied, Bratislava
[5] CHALOUPKOVÁ, Z. (1991). Současnost a romská populace. Výskumný ústav práce a sociálnych vecí, Bratislava
[6] KVAPILOVÁ, E. (2000). Rôznorodosť rodinných foriem - výzva pre sociálnu politiku. Sociológia 32, č. 5
[7] LENCZOVÁ, M. (1996). Účinnosť verejnej politiky rodiny – Výskum realizácie rodinných funkcií v rómskych rodinách. Výskumný ústav práce, sociálnych vecí a rodiny. Bratislava
[8] MENTEL, A. (2003). Sociálne a kultúrne faktory ovplyvňujúce variabilitu mimomanželskej plodnosti. In: Rodina. 9. slovenská demografická konferencia, Tajov 17. – 19. 9. 2003. Zborník vedeckých prác. SŠDS Bratislava
[9] MLÁDEK, J. (2003). Kohabitácie a kohabitanti na Slovensku – národnostná, religiózna, vzdelanostná a priestorová štruktúra. In: Rodina. 9. slovenská demografická konferencia, Tajov 17. – 19. 9. 2003. Zborník vedeckých prác. SŠDS Bratislava
[10] MLÁDEK, J., ŠIROČKOVÁ, J. (2003). Demografické aspekty rodinného správania. Prezentácia na Česko - poľsko - slovenskom seminári. Karlova univerzita, Praha
[11] MLÁDEK, J., ŠIROČKOVÁ, J. (2004). Štúdie. Kohabitácie ako jedna z foriem partnerského spolužitia obyvateľstva Slovenska. Sociológia 36, č.5

[12] ŠIROČKOVÁ, J. (2003). Zmeny a nové formy rodinného správania. In: Rodina. 9. slovenská demografická konferencia, Tajov 17. – 19. 9. 2003. Zborník vedeckých prác. SŠDS Bratislava
[13] ŠÚ SR. Databáza ŠÚ SR - SĽDB 1991, SODB 2001

BC. TATIANA LACHOVÁ
Prírodovedecká fakulta UK, Katedra humánnej geografie a demogeografie, 5. ročník
taila@zoznam.sk

Etnická štruktúra okresov SR podľa sčítania v roku 1880

Juraj Majo

Úvod

Sčítanie obyvateľstva v roku 1880 bolo už druhým sčítaním v Rakúsku-Uhorsku v ére modernej štatistiky. Bolo však prvým, z ktorého je možné získať údaje o etnických pomeroch na našom území, avšak len prostredníctvom uvedenia materinského jazyka a nie na základe deklaratívnosti. Údaje o materinskom jazyku a religiozite sú publikované na úrovni obcí v publikácií A magyar korona országaiban az 1881.évi népszámlálás, föbb eredményei. Vzhľadom na historický, geografický a demografický význam týchto údajov, sa etnickou štruktúrou z tohto sčítania zaoberali viacerí autori či už prostredníctvom vedeckých článkov (Žudel, 1991, Mazúr 1974), resp. kartografických interpretácií jeho údajov (Benža, 2002). J. Žudel analyzuje sčítanie z tohto roku v pôvodnej administratívnej štruktúre, podobne ako M. Benža v kartografickej interpretácií v Atlase krajiny SR. E. Mazúr údaje zrekonštruoval do administratívnej štruktúry okresov zo 70. rokov a porovnával s viacerými sčítaniami. V tomto príspevku analyzujeme etnickú štruktúru z roku 1880 v administratívnej štruktúre okresov platnej od roku 1996. Analyzovali sme údaje za 4 najpočetnejšie etniká na území Slovenska. Zvyšné, málopočetné etniká sú zahrnuté pod kategóriu ostatné.

Slováci

Najpočetnejším etnikom na území dnešného Slovenska boli v roku 1880 Slováci s podielom 61,15% obyvateľov. Relatívne hodnoty výskytu slovenskej národnosti narastali od juhu smerom na sever a od východu smerom na západ. Najvyšší podiel slovenskej národnosti mali okresy na severe a severozápade Slovenska, pôvodne severné časti Trenčianskej stolice a Oravská stolica, tvoriace dnes okresy Žilinského kraja. Hodnoty v týchto okresoch presahovali 90%. (Kysucké Nové Mesto 93,88%, Dolný Kubín 92,09%, Námestovo 93,45%, a pod.). Podobne vysoké hodnoty zastúpenia slovenskej národnosti sa vyskytujú aj v Banskobystrickom kraji, v okresoch Krupina (91%), Brezno (92,98%), Poltár (87,13%). Najvyššie hodnoty na celom území Slovenska dosiahol okres Detva s podielom 94,25% obyvateľstva hlásiacich sa k slovenskej národnosti. Na západnom Slovensku sa vysoké hodnoty vyskytujú aj v okresoch Myjava (91,65%), Bánovce nad Bebravou (90,24%), Senica (88,15%) a pod. Na východnom Slovensku mali najvyššie hodnoty okresy Košice III. (93,06%), Košice II. (86,37%), Sabinov (85,03%). Najnižšie hodnoty podielu slovenskej národnosti sa vyskytovali na území, ktoré prevažne obývala maďarská národnosť a to najmä okresy juhozápadného Slovenska, kde v niektorých okresoch nedosahovala hodnota relatívneho podielu obyvateľov slovenskej národnosti ani 5%. Ide o okresy Bratislava II. (4,23%), Bratislava V. (2,34%), Komárno (3,85%). Percentuálne najmenej Slovákov vôbec, žilo v okrese Dunajská Streda, kde podiel dosiahol iba 1,07 %. Minimálne zastúpenie mala slovenská národnosť aj v okrese Medzilaborce, s podielom 3,31%. Tento však na rozdiel od predošlých obývalo prevažne obyvateľstvo hlásiace sa k rusínskej národnosti.

Maďari

Druhým najpočetnejším etnikom s pomerne výrazným zastúpením boli Maďari s podielom 22,21%. Ich výskyt je prirodzene viazaný na južné okresy Slovenska a pomerne silné zastúpenie mali aj v mestách na celom území Slovenska. Najsilnejšie postavenie mala maďarská národnosť na juhozápade Slovenska, kde konkrétne v okresoch Dunajská Streda a Komárno mala podiel viac ako 90 %, (Dunajská Streda 92,58%, Komárno 90,03%). Výrazné zastúpenie bolo aj v okresoch Šaľa (74,85%), resp. Levice (66,52%) a dnešný mestský okres Bratislava II. (66,11%). V južných okresoch stredného a východného Slovenska tento podiel

nedosahoval vysoké hodnoty západného Slovenska, avšak pohyboval sa okolo 50% podielu. Na strednom Slovensku sú to okresy Rimavská Sobota (62,32%), Lučenec (49,41%), Rožňava (48,24%) Veľký Krtíš (45,05%). Ešte nižšie hodnoty dosiahli južné okresy východného Slovenska, kde iba okres Trebišov mal miernu prevahu maďarského obyvateľstva nad ostatným (50,4%), výraznejšie zastúpenie bolo aj v okresoch Košice – okolie (36,38%) a v meste Košice, resp. okrese Košice I. (35,13%). Celkovo najnižší podiel maďarského obyvateľstva dosahovali okresy severného a najmä severozápadného a západného Slovenska – Námestovo (0,17%), Stará Ľubovňa (0,55%), Kysucké Nové Mesto (0,39%), Skalica (0,67%) a pod. Veľmi nízke hodnoty mali aj tie okresy ležiace v strede a na východe Slovenska, ktoré mali tiež vysoký podiel slovenskej národnosti (Detva – 0,85%, alebo Košice – 0,28% a Brezno – 1,30%).

Nemci

Nemecké obyvateľstvo dosiahlo pri sčítaní obyvateľstva v roku 1880 podiel 9,28% a bola tretím najpočetnejším etnikom na území Slovenska. Jeho výskyt nie je viazaný na jednu oblasť, ale dosiahlo pomerne rovnomerné rozmiestnenie po celom území, významnejšie zastúpenie malo najmä však v mestách a vo viacerých okresoch Hornej Nitry, Spiša a okolí Bratislavy. Najsilnejšie zastúpenie mala nemecká národnosť v Bratislave, kde dosahovala 63,41% a bola tak najpočetnejšou národnosťou v meste. Silné zastúpenie mala aj v okolitých okresoch, ktoré dnes tvoria mestské časti Bratislavy. Najsilnejšou v tejto oblasti bola v piatom bratislavskom okrese, kde dosahovala až 59,12% a podobne ako v prvom bratislavskom okrese, aj tu bola najviac zastúpenou národnosťou. V ostatných okresoch dnešnej Bratislavy dosiahla pomerne významné zastúpenie v okrese Bratislava II. (24,82%) a Bratislava IV. (20,55%). Ďalšou oblasťou s výrazným zastúpením bola oblasť Spiša, kde v okrese Kežmarok dosiahla nadpolovičné zastúpenie (50,23%) a v okrese Gelnica bola s podielom 46,65% tiež najpočetnejšou národnosťou v okrese. Z ostatných okresov mala nemecká národnosť silné zastúpenie v okrese Turčianske Teplice (35,82%), Žiar nad Hronom (32,8%) a Prievidza (29,47%).

Rusíni

Špecifické postavenie v rámci Slovenska mali Rusíni, ktorých zastúpenie bolo takmer výlučne viazané na okresy východného Slovenska. Nadpolovičný podiel dosiahli pritom iba v dvoch okresoch Slovenska – Medzilaborce (80,45%) a Svidník (54,27%). Blízko pod hranicou 50% sa podiel obyvateľov rusínskej národnosti pohyboval aj v okresoch Snina (46,22%), Stropkov (45,94%) a Stará Ľubovňa (40,44%). Zastúpenie v ostatných okresoch Slovenska dosiahlo takmer výlučne nulový podiel medzi národnosťami a na celom území SR dosiahla rusínska národnosť iba 3,09%.

Ostatní

Kategória ostatných národností tvorí sumu málopočetných národností zastúpených na Slovensku. Najpočetnejšou z týchto národností bola chorvátska národnosť, ktorá mala najsilnejšie zastúpenie v okresoch Bratislava V. (26,08%) a Bratislava IV (19,74%).

Záver

Problematika poznávania etnickej štruktúry Slovenska je nesmierna rozsiahla a geografický aspekt je len jedným z mnohých, pričom práve priestorová analýza a interpretácia jej údajov predstavuje jeden z najnázornejších. Prostredníctvom zrekonštruovaných a zdigitalizovaných dát sa v prostredí geoinformačných technológií vytvára jedinečná možnosť ich variabilnej kartografickej interpretácie a komparácie so súčasnými údajmi.

Literatúra:

A magyar korona országaiban az 1881.év elején végrehajtott népszámlálás. II. kötet. Budapest, 1882

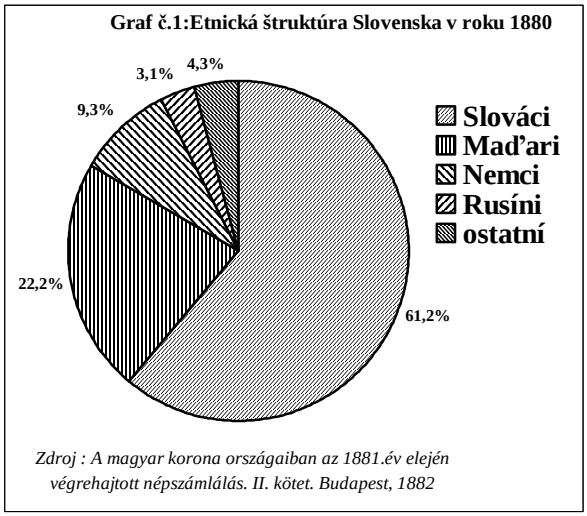
BENŽA, M. (2002). Národnostná štruktúra v roku 1880. In: *Atlas krajiny Slovenskej republiky*. Bratislava, Banská Bystrica (MŽP SR a SAŽP) s. 157

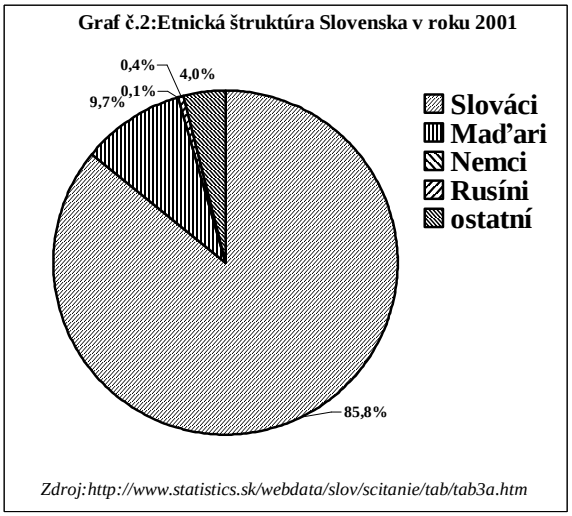
MAZÚR, E. (1974). Národnostné zloženie. In: Slovensko – Ľud I.. Bratislava (Obzor) s. 440 - 451

ŽUDEL, J. (1993). Národnostná štruktúra obyvateľstva v roku 1880- In: *Geografický časopis* č.1, roč 45 (Veda) s.3-17

http://www.statistics.sk/webdata/slov/scitanie/tab/tab3a.htm

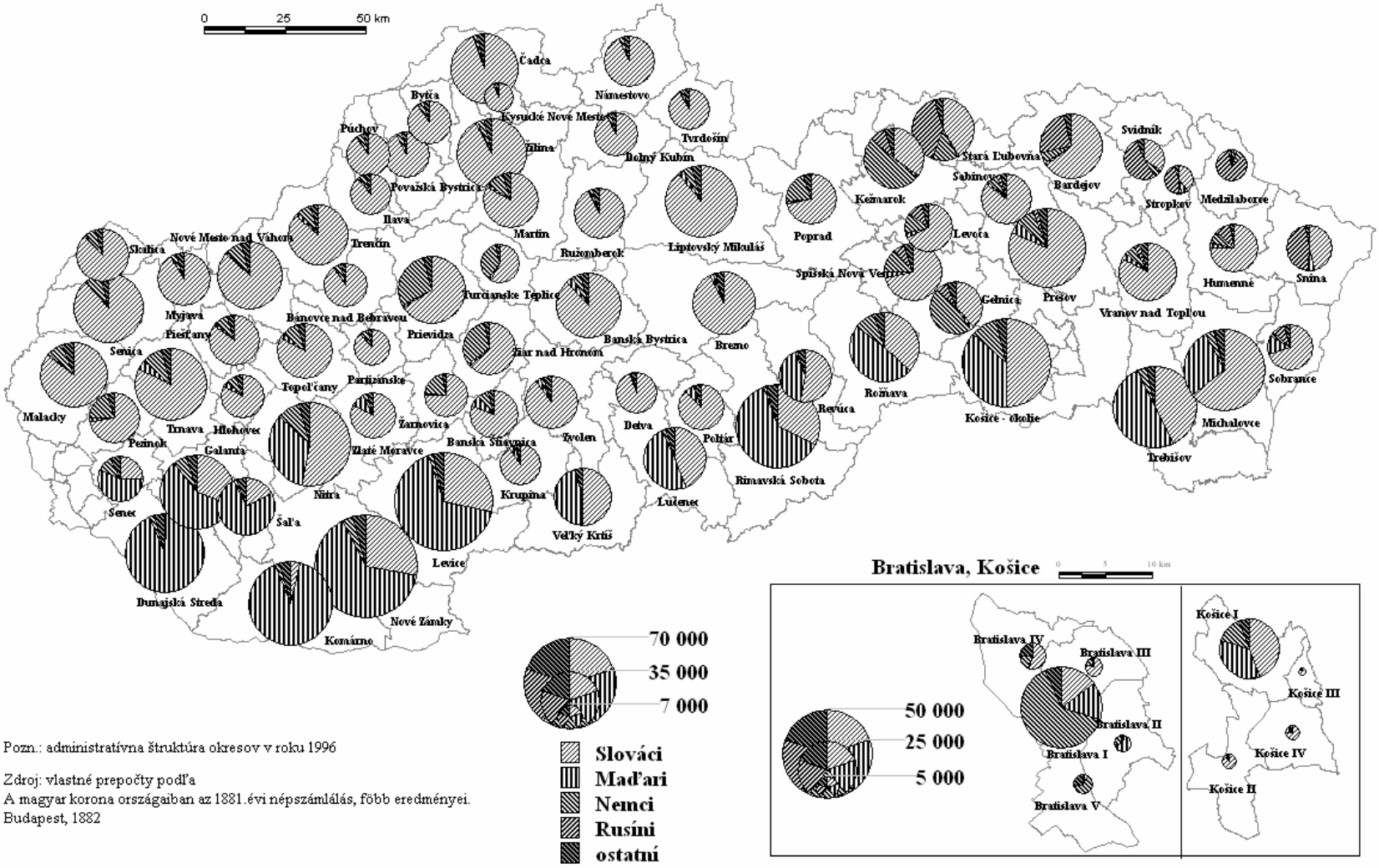
Bc. Juraj Majo
Univerzita Komenského, Prírodovedecká fakulta, 4. ročník
jurm@stonline.sk





Mapa č.1: Etnická štruktúra obyvateľstva okresov SR podľa sčítania z roku 1880

autor Juraj Majo



Pozn.: administratívna štruktúra okresov v roku 1996

Zdroj: vlastné prepočty podľa
A magyar korona országaiban az 1881.évi népszámlálás, főbb eredményei.
Budapest, 1882

MANŽELSKÁ PLODNOSŤ PODĽA DĹŽKY TRVANIA MANŽELSTVA A PODĽA PORADIA NARODENÉHO DIEŤAŤA

Barbora Marônková

Úvod

Po druhej svetovej vojne možno v mnohých krajinách Európy pozorovať zmeny v demografickom správaní obyvateľov a meniaci sa postoj najmä k uzatváraniu manželstiev a s ním spojenou plodnosťou vydatých žien. Napriek tomu je ešte stále v Slovenskej republike manželská plodnosť dominantná, keďže až 80 % detí sa rodí v manželstve. Vývoj pôrodnosti a plodnosti súvisí na Slovensku s postavením rodiny. Dôraz na rodinu ako základ štátu bol v minulosti nielen krédom, ale sa v podstate zhodoval i s vnútornými postojmi a reálnym správaním obyvateľov (Rychtaříková, J. 1996). Založenie vlastnej rodiny bolo jednou z mála možností sebarealizácie a osamostatnenia mladých ľudí v tom čase.

Uvoľnenie politického systému a zavedenie demokratických prístupov paradoxne prispelo k uvoľneniu rodinných foriem a správania. Po roku 1989 dochádza k výrazným dynamickým zmenám v manželskej plodnosti. Možno pozorovať súbežné javy, kým intenzita manželskej plodnosti klesá, intenzita mimomanželskej plodnosti narastá. Vzhľadom na tieto skutočnosti si problematika manželskej plodnosti zaslúži väčšiu pozornosť a podrobnejšiu analýzu.

Cieľom príspevku je analýza manželskej plodnosti podľa dĺžky trvania manželstva a podľa poradia narodeného dieťaťa, čo je problematika menej analyzovaná v odbornej literatúre i v súvislosti s veľkými nárokmi, ktoré kladie na dátovú základňu. Ako sa mení manželská plodnosť na Slovensku, a aká je pravdepodobnosť zväčšovania rodiny na Slovensku? Môžeme pozorovať nejaké zmeny?

Výpočet redukovaných mier manželskej plodnosti a pravdepodobnosti zväčšovania rodiny

Pri analýze manželskej plodnosti je možné sledovať zmeny manželskej plodnosti podľa veku matky (resp. otca), dĺžky trvania manželstva a poradia narodeného dieťaťa.

Na výpočet tabuliek manželskej plodnosti podľa dĺžky trvania manželstva je vhodné použiť charakteristiku - redukované miery manželskej plodnosti. Podľa spôsobu triedenia narodených detí v manželstve možno rozlišovať prípad, keď:

a. manželsky narodení sú triedení podľa roku sobáša a dĺžky trvania manželstva,
b. manželsky narodení sú triedení podľa dĺžky trvania manželstva.

Pri výpočte tabuliek manželskej plodnosti boli manželsky narodení triedení práve podľa dĺžky trvania manželstva.

Redukované miery manželskej plodnosti, kde manželsky narodení (všetci narodení) sú triedení podľa doby uplynulej od sobáša, možno vyjadriť vzorcom

$$f_x^{m.r} = \frac{{}_{t}^{z-1,z}N_x^m}{\frac{{}_{t-x}S + {}_{t-x-1}S}{2}} * 1000,$$

kde v čitateli vystupujú narodení v *x* dokončených rokoch trvania manželstva v kalendárnom roku *t* (t.j. môžu byť z dvoch sobášnych kohort z-1; z), v menovateli je priemer dvoch odpovedajúcich sobášnych kohort.

Sumou redukovaných mier manželskej plodnosti, rátaných pre určité kalendárne obdobie, dostaneme hodnotu úhrnnej manželskej plodnosti, ktorú možno vyjadriť aj týmto zápisom

$$úp = \sum_{x=0}^{21+} f_x^{m,r},$$

alebo konečnú manželskú plodnosť v prípade tabuliek pre sobášnu kohortu.

Tabuľky manželskej plodnosti podľa poradia narodeného dieťaťa sú založené na výpočte pravdepodobností narodenia dieťaťa i-tého poradia.

Pravdepodobnosť zväčšovania rodiny udáva, s akou pravdepodobnosťou bude mať žena s n deťmi n+1 detí.

V tabuľkách 1. poradia je časovou premennou doba od sobáša po narodenie prvého dieťaťa, študovanými udalosťami sú pôrody 1. poradia, ktorými ženy, resp. manželské páry vystupujú zo súboru exponovaných. V tabuľkách 2. poradia je časovou premennou doba od narodenia prvého dieťaťa, študovanými udalosťami sú pôrody 2. poradia a exponovanou populáciou súbor žien, resp. manželských párov, ktorým sa narodilo dieťa 1. poradia. Narodením druhého dieťaťa, žena, resp. pár opúšťa pole pozorovania.

Všeobecne tabuľky manželskej plodnosti i- tého poradia popisujú rád rodenia i- tých detí, súborom exponovaných je populácia majúca (i-1) detí a časovou premennou je doba medzi narodením dieťaťa (i-1)-ho poradia a i-tého poradia.

Pri výpočte pravdepodobnosti a_0 sa narodení 1. poradia podľa doby od sobáša vzťahujú k počtu sobášov, ku ktorým príslušia (t.j. berieme priemer sobášov z dvoch susedných rokov):

$$a_0 = \sum_{x=0}^{w-1} \frac{{}_tN_x^1}{\frac{{}_{t-x}S + {}_{t-x-1}S}{2}},$$

kde *w* je dĺžka trvania manželstva, kedy pokladáme plodnosť za ukončenú; N^1 predstavuje narodených v manželstve 1. poradia; *t* predstavuje kalendárny rok, *x* predstavuje pri pravdepodobnosti a_0 prvopôrodný interval, t.j. doba od sobáša po 1. pôrod, kým pri pravdepodobnosti a_1, a_2, atď. *x* predstavuje medzipôrodný interval, t.j. doba medzi narodením dieťaťa (i-1)-ho poradia a i-tého poradia.

Pravdepodobnosť a_1 počítame rovnakým spôsobom, len v menovateli neuvažujeme sobáše, ale berieme priemer manželsky narodených 1. poradia z dvoch susedných rokov a v čitateli sú manželsky narodení 2. poradia ($_tN^2$):

$$a_1 = \sum_{x=0}^{w-1} \frac{{}_tN_x^2}{\frac{{}_{t-x}N^1 + {}_{t-x-1}N^1}{2}},$$

analogicky počítame pravdepodobnosť a_2 , kde v menovateli je priemer manželsky narodených 2. poradia z dvoch susedných rokov a v čitateli sú narodení v manželstve 3. poradia ($_tN^3$):

$$a_2 = \sum_{x=0}^{w-1} \frac{{}_tN_x^3}{\frac{{}_{t-x}N^2 + {}_{t-x-1}N^2}{2}}$$

(Pavlík, Z., Rychtaříková, J., Šubrtová, A. 1986, s.307-322).

Vývoj redukovaných mier manželskej plodnosti na Slovensku v rokoch 1950-2002

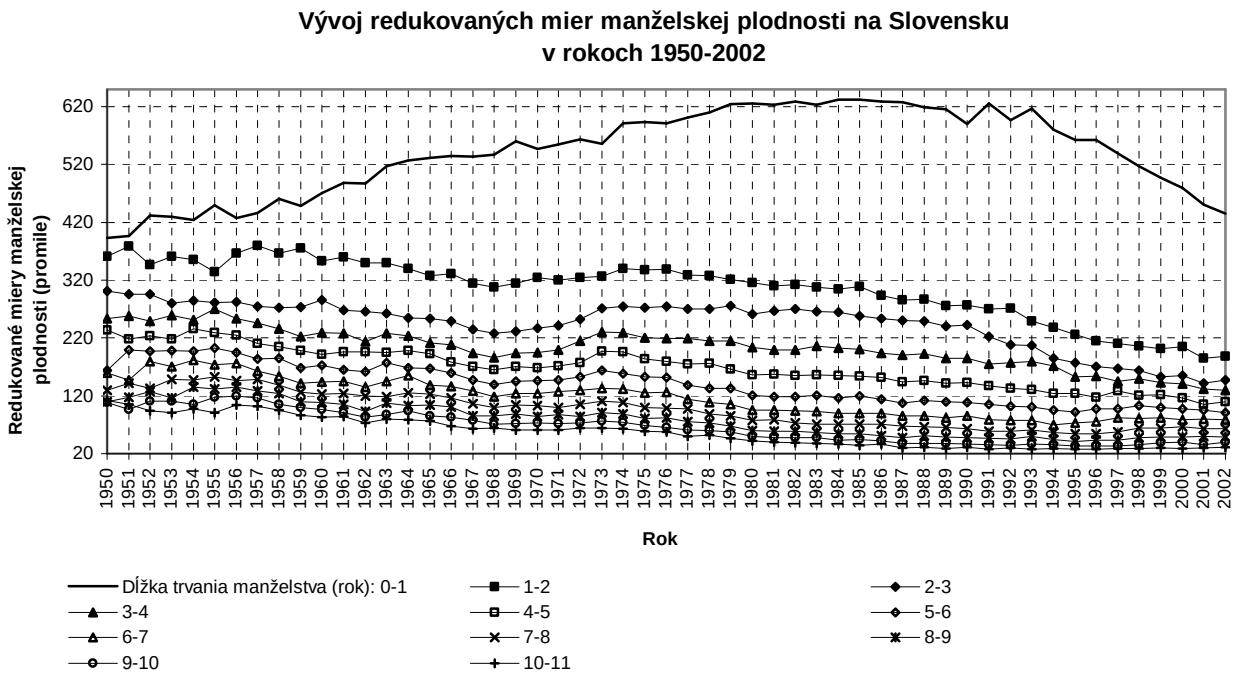
Na prvý pohľad je zrejmé (graf č.1), že v sledovanom období dochádzalo ku zmenám úrovne manželskej plodnosti podľa doby uplynulej od sobáša. V priebehu päťdesiatych až do prvej polovice osemdesiatych rokov 20. storočia dochádza k zvyšovaniu úrovne hodnôt tohto ukazovateľa pri dĺžke trvania manželstva do jedného roka, keď v roku 1950 predstavoval 392,88 ‰ a v roku 1984 už dosiahol úroveň 632,48 ‰, ktorá zároveň predstavovala najvyššiu úroveň redukovanej miery manželskej plodnosti v sledovanom období.

Od druhej polovice osemdesiatych rokov 20. storočia pozorujeme zastavenie tohto trendu a v deväťdesiatych rokoch 20. storočia, najmä v ich druhej polovici sledujeme trend opačný, t.j. pokles redukovanej miery manželskej plodnosti.

Pri sledovaní redukovaných mier manželskej plodnosti pri dĺžke trvania manželstva 1-2 roky možno zaznamenať výrazne nižšie hodnoty ako pri tomto ukazovateli pri dĺžke trvania manželstva do jedného roku. Takýto priebeh plodnosti nasvedčuje tomu, že páry, ktoré uzatvorili manželstvo, hneď realizujú svoju plodnosť, príp. bolo tehotenstvo budúcej nevesty bezprostredným impulzom k uzatvoreniu manželstva.

Analýza redukovanej miery manželskej plodnosti pri dĺžke trvania manželstva 1-2 roky nasvedčuje, že výrazne klesá, keď na začiatku nami sledovaného obdobia predstavovala 361,29 ‰ (r.1950) a v roku 2002 už len 188,18 ‰, čo len dokazuje neustály pokles manželskej plodnosti.

Graf č.1



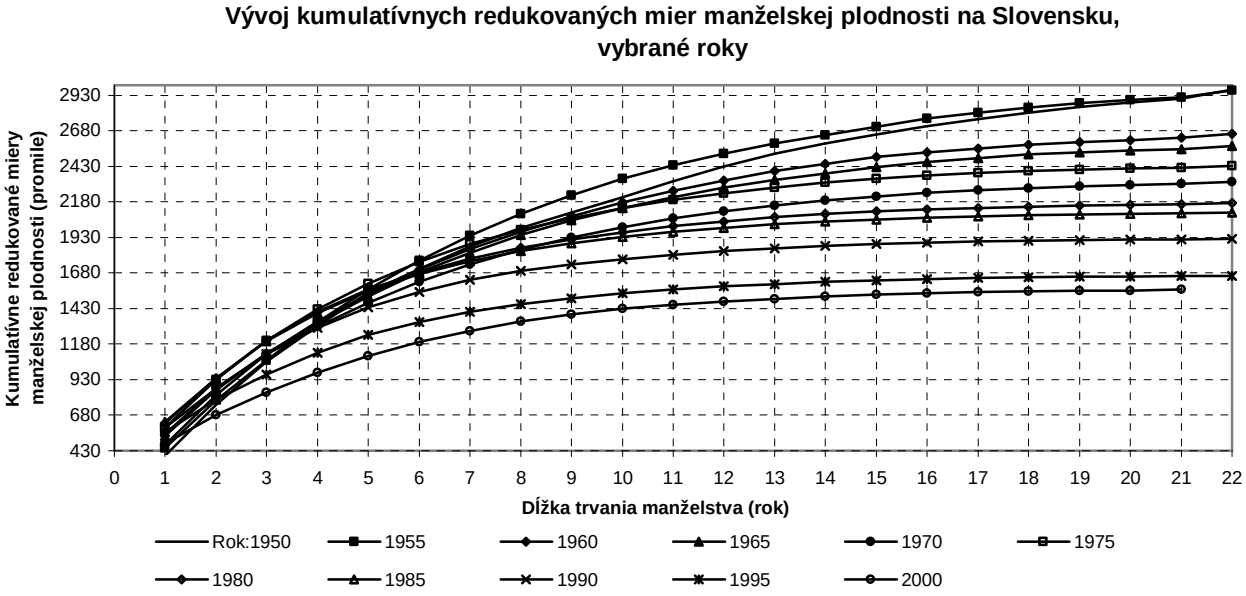
Zlom vo vývoji redukovaných mier manželskej plodnosti možno pozorovať od druhej polovice osemdesiatych rokov 20. storočia, ale predovšetkým v rokoch deväťdesiatych. Je dôsledkom všeobecného trendu znižovania plodnosti, a taktiež svedčí o zmene demografického správania v tom význame, že tehotenstvo už nie je považované za tak veľmi závažný dôvod k urýchlenému uzatvoreniu manželstva (Fiala, T. 2001).

Z grafu kumulatívnych redukovaných mier manželskej plodnosti možno opäť pozorovať (graf č.2), že plodnosť slovenských žien v priebehu rokov 1950-2002 klesá. Napríklad pri dĺžke trvania manželstva 10 rokov, dosiahol tento ukazovateľ v roku 1955 úroveň 2345,02 ‰, o 30 rokov neskôr úroveň 1932,51 ‰ a v roku 2002 úroveň 1360,98 ‰. Pri dlhšom trvaní manželstva sa rozdiel medzi hodnotami ukazovateľa v pozorovanom časovom rade ešte viac prehlbuje. Menšie zvýšenie plodnosti vydatých žien možno pozorovať v roku 1975, ktoré možno vysvetliť ako dôsledok pronatalitných opatrení, a taktiež tým, že silné ročníky z päťdesiatych rokov 20. storočia vstúpili do fertilného veku.

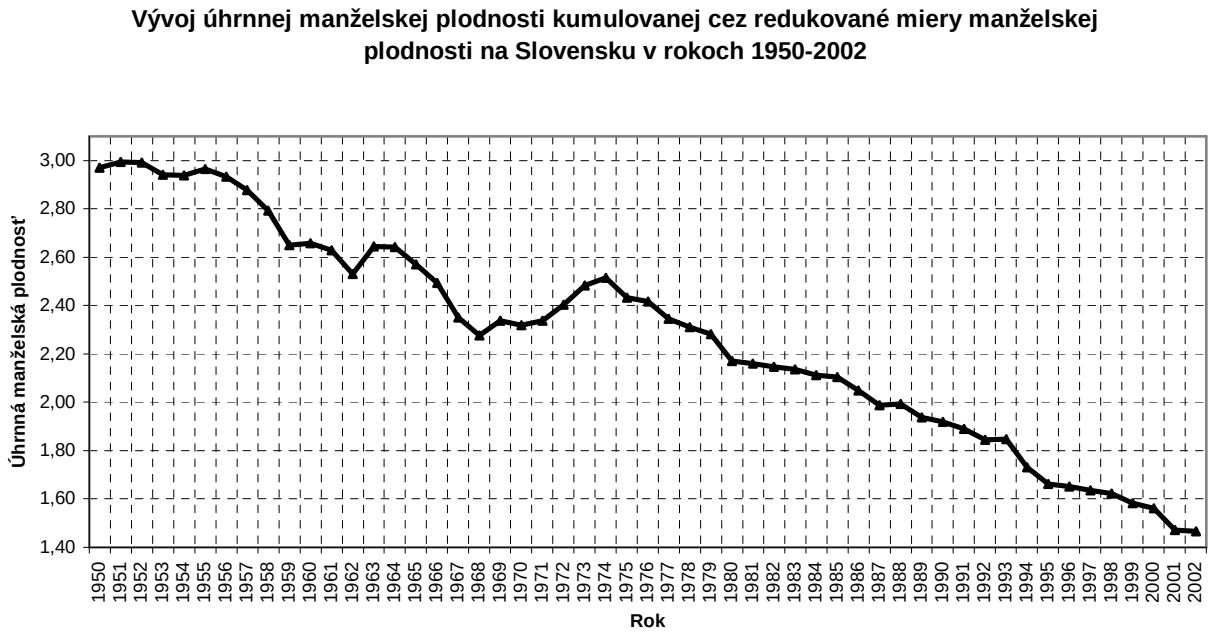
Úhrnná manželská plodnosť kumulovaná cez redukované miery manželskej plodnosti (graf č.3) dosahuje od roku 1950 až 2002 čoraz nižšie hodnoty. Vo vývoji sa dá určiť jeden výraznejší a dva menšie vrcholy. Prvý je vidieť v päťdesiatych rokoch 20. storočia (r.1950-1955), druhý hodnotami nižšími od prvého vrcholu sa dosiahol v rokoch 1963-1964 a tretí vrchol v roku 1974. Tieto obdobia súvisia so zvýšenou pôrodnosťou. Obdobie päťdesiatych rokov 20. storočia možno považovať za kompenzačnú fázu po druhej svetovej vojne. Hlavným dôvodom nárastu zaznamenaného v prvej polovici sedemdesiatych rokov môže byť skutočnosť, že silné povojnové ročníky vstúpili do fertilného veku, a taktiež štát robil v tom čase rôzne pronatalitné opatrenia.

Od roku 1974 je badateľný sústavný pokles úhrnnej plodnosti vydatých žien na Slovensku, výnimku tvorí iba rok 1988 (1,992), kedy oproti predchádzajúcemu roku úhrnná plodnosť narástla (1,988 – r.1987). Aj vývoj tohto ukazovateľa dokumentuje pokles manželskej plodnosti, ktorý je citeľnejší od sedemdesiatych rokov 20. storočia a zintenzívnil sa v deväťdesiatych rokoch. Príčinou prepadu úhrnnej manželskej plodnosti na tak nízku úroveň môže byť najmä odkladanie pôrodov do vyššieho veku žien. Otázkou zostáva, či sa odložená reprodukcia vo vyššom veku zrealizuje úplne, alebo iba čiastočne. Predpokladá sa, že odložené pôrody sa realizujú len čiastočne, keďže mnoho mladých rodín si v prípade zmeny svojich životných podmienok môže prehodnotiť svoje predchádzajúce rozhodnutie o nutnosti ďalšieho plánovaného dieťaťa, resp. sa môže rozhodnúť pre bezdetné manželstvo alebo bezdetné partnerské spolužitie (Vaňo, B. 2000).

Graf č.2



Graf č.3



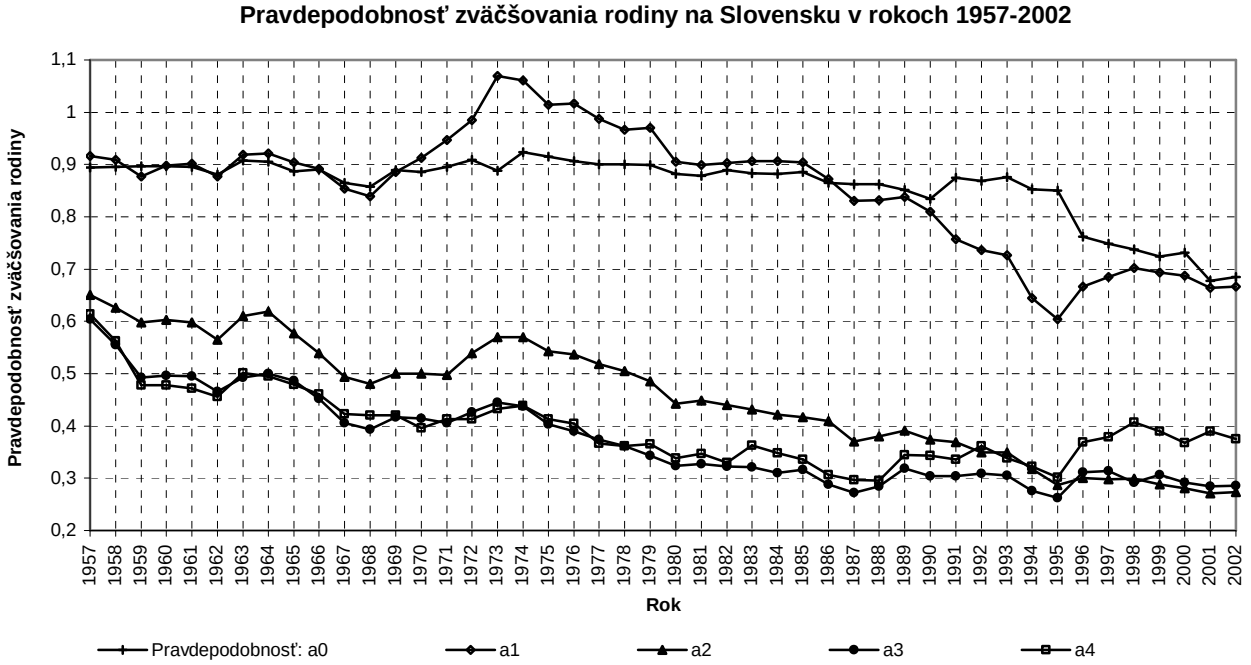
Pravdepodobnosť zväčšovania rodiny na Slovensku v rokoch 1957-2002

Pri komparácií pravdepodobností a_0 a a_1 (graf č.4) sledujeme takmer rovnaký priebeh vývoja v rokoch 1957-1969. V nasledujúcom období až do roku 1986 nadobúda pravdepodobnosť a_1 vyššie hodnoty. To znamená, že v tomto období bola vyššia pravdepodobnosť, že sa v rodine narodí dieťa druhého poradia ako pravdepodobnosť, že sa v rodine narodí dieťa, t.j. prechod od bezdetnosti k narodeniu dieťaťa prvého poradia. Avšak od roku 1986 je pravdepodobnosť, že sa narodí v manželstve prvé dieťa vyššia, ako pravdepodobnosť a_1 (narodenie dieťaťa druhého poradia). Od spomínaného roku zároveň hodnoty obidvoch pravdepodobností klesajú.

Pravdepodobnosti a_2, a_3, a_4 dosahujú v celom pozorovanom vývoji nižšie hodnoty ako pravdepodobnosti a_0 a a_1. Krivky znázorňujúce priebeh pravdepodobností a_3 a a_4 majú v období 1957-2002 takmer identický priebeh. Výraznejšie sa začínajú líšiť od začiatku deväťdesiatych rokov 20. storočia, keď pravdepodobnosť, že sa v manželstve narodí päť detí dosahuje vyššie hodnoty nad

pravdepodobnosťou narodenia dieťaťa štvrtého poradia. Všetky vypočítané pravdepodobnosti však od začiatku až ku koncu sledovaného obdobia poklesli.

Graf č.4



V súčasnosti je vyššia pravdepodobnosť, že sa narodí v manželstve jedno dieťa, ako keby sa v ňom mali narodiť dve a viac detí. Na Slovensku sa tak v posledných rokoch začína presadzovať model rodiny s jedným dieťaťom ako výsledok dlhodobého spoločenského vývoja a transformačných procesov po novembri 1989, ktoré vyvolali určité zmeny, najmä v ekonomickom a spoločenskom postavení mladých ľudí.

Záver

Výrazne sa zmenilo správanie manželských párov. Všeobecne sa dá tvrdiť, že hodnoty redukovaných mier manželskej plodnosti a hodnoty pravdepodobností poklesli v sledovanom období. Výnimku tvorí redukovaná miera manželskej plodnosti pri dĺžke trvania manželstva do jedného roku, pri ktorej úroveň hodnôt v roku 2002 oproti roku 1950 narástla. Počas celého vývoja sa najviac detí rodí do jedného roku trvania manželstva, avšak oproti sedemdesiatym a osemdesiatym rokom 20. storočia možno v deväťdesiatych rokoch zaznamenať pokles počtu narodených detí. Súčasne sa mení aj tempo tohto poklesu, ktoré sa najintenzívnejšie prejavuje v druhej polovici deväťdesiatych rokov 20.storočia. Možno predpokladať, že žien, resp. párov, resp. manželstiev bez detí v budúcnosti narastie a súčasne sa zníži i úroveň sobášnosti (Rychtaříková, J 1996).

Úhrnná manželská plodnosť sa rapídne znížila, keď v päťdesiatych rokoch sa priemerne narodilo jednej žene v manželstve 2,9 detí, avšak už v roku 2002 ani nie 1,5 detí, čo predstavuje výraznú zmenu v intenzite manželskej plodnosti.

Na Slovensku sa začína postupne preferovať model rodiny s jedným dieťaťom. Takéto správanie možno vysvetliť sociálno-ekonomickými faktormi, keď znížená plodnosť sa spája s úpadkom ekonomickej situácie v deväťdesiatych rokoch 20. storočia; idividuálnymi faktormi, kedy môžeme pozorovať nárast individualizácie a sekularizácie, emancipácie, zmeny v hodnotovej orientácií ľudí a dieťa je považované za ekonomickú zátaž, ale taktiež možno vysvetliť znižovanie pôrodnosti a plodnosti faktormi demografickými, medzi ktoré možno zaradiť rast priemerného veku pri pôrode a pri prvom pôrode, s čím súvisí odkladanie pôrodov do vyššieho veku a zostáva otázne, či sa odložená reprodukcia vo vyššom veku zrealizuje čiastočne, alebo iba úplne.

Okrem toho, že sa následky poklesu pôrodov prejavujú vo vekovej štruktúre populácie Slovenska, výrazné zníženie počtu rodiacich sa detí bude mať výrazný vplyv na školský systém a prejaví sa vo všetkých etapách života týchto generácií.

Použitá literatúra a pramene

Fiala, T. (2001): Vývoj manželské plodnosti prvního pořadí v České republice během posledních padesáti let. Demografie, Roč. 45, Č. 2, 93-110.
Infostat- Výskumné demografické centrum (2001): Obyvateľstvo Slovenska 1945-2000. Bratislava.
Pavlík, Z., Rychtaříková, J., Šubrtová, A. (1986): Základy demografie. 1. vyd. Praha, ČSAV, 736 s.
Státní/Federální statistický úřad: Pohyb obyvateľstva v ČSSR/ČSFR/SSR 1950-1990. Praha, vlastné výpočty.
Štatistický úrad Slovenskej republiky: Pohyb obyvateľstva v Slovenskej republike 1991-2002. Bratislava, vlastné výpočty.
Rychtaříková, J. (1993): Současné demografické postavení Českých zemí a Slovenska v Evropě. Demografie, r.35, č.1, s.20-23.
Rychtaříková, J. (1996): Současné změny charakteru reprodukce v České republice a medzinárodní situace. Demografie, roč. 38, č. 2, s. 77-89.
Vaňo, B. (ed.) (2000): Populačný vývoj v SR 1999. Bratislava, Infostat- Výskumné demografické centrum.

Bc. Barbora Marônková
Katedra humánnej geografie a demogeografie, Univerzita Komenského v Bratislave, Prírodovedecká fakulta, Mlynská dolina, 842 15 Bratislava 4
e-mail: bmaronkova@post.sk

Systém riadenia investičného rizika do portfólia cenných papierov
(podpora rozhodovania portfóliového managementu s využitím systému SAS)

Branislav Rajčáni

Úvod

Riadenie štátneho dlhu a likvidity je nezbytnou súčasťou každého štátu. V roku 2005 je splatných 177 mld. Sk štátneho dlhu Slovenskej republiky. Štát zabezpečuje svoju likviditu aj prostredníctvom emisií štátnych obligácií. Štát sa tak fakticky stáva dlžníkom. V mojej práci som sa zameral na problematiku výnosnosti štátnych obligácií s ročnou kupónovou sadzbou, avšak s rôznou lehotou splatnosti. Chcem dokázať koreláciu medzi výnosnosťami jednotlivých sledovaných emisií. Výnosnosť je jeden z dôležitých faktorom pri rozhodovaní sa portfóliového managera o investícii do cenných papierov. Zároveň chcem poukázať na závislosť medzi vývojom výnosnosti a vývojom Slovenského dlhopisového indexu SDX pre štátne dlhopisy.

Finančný trh

Finančný trh je miesto, kde sa stretáva ponuka a dopyt. Miesto, kde sa stretávajú tí, ktorí majú nedostatok finančných prostriedkov a zároveň tí, ktorí vlastnia voľné finančné prostriedky v podobe úspor rôznych ekonomických subjektov a hľadajú spôsoby a formy ich zhodnotenia. Investor zároveň musí riešiť problematiku objemu voľných finančných prostriedkov, potreby vlastnej likvidity, likvidity nástroja finančného trhu, do ktorého hodlá investovať ako aj mieru rizikovosti a s tým súvisiacej výnosnosti investičného nástroja. Z praxe platí, že medzi výnosom a mierov rizika platá priama úmera, teda vyššie riziko predpokladá aj vyšší výnos a opačne. Na finančnom trhu realizuje svoje záujmy aj štát.

Obligácie

Z terminologického hľadiska je obligácia presnejším pojmom ako dlhopis. Obligácie sú súčasťou kapitálových trhov, kde sa obchoduje s dlhodobými investíciami. Obligácia je cenný papier, ktorý vyjadruje záväzok emitenta(dlžníka) voči vlastníkovi cenného papiera. Obligácia teda vyjadruje úverový vzťah medzi dlžníkom a veriteľom stanovujúcim lehotu splatnosti, výšku úrokov a spôsob platby istiny a úrokov, avšak vlastník obligácie nedisponuje právom riadiť emitenta obligácie. Pre ďalšiu pozorovania som si zvolil obligácie s fixnou(stálou) kupónovou sadzbou s vyplatením menovitej hodnoty v deň splatnosti.

Agentúra pre riadenie dlhu a likvidity(ARDAL) zabezpečuje optimalizáciu štruktúry štátneho dlhu SR a náklady s ním spojené na základe analýzy trhu, analýzy portfólia trhu a vypracovaného systému riadenia rizík. V spolupráci s MF SR sa snaží o zvýšenie likvidity štátnych cenných papierov. Za najefektívnejšiu formu zadlžovania štátu tak pokladajú emitovanie štátnych obligácií na domácom a zahraničnom trhu. Je to zároveň dôležitá súčasť pri integrácii finančného trhu a finančného riadenia okruhu verejných financií s krajinami EU. V kompetencii ARDAL-u je aj riadenie emisií štátnych obligácií. Tieto sa stávajú vlastníctvom veriteľov, teda tých, ktorí požičiavajú svoje voľné finančné prostriedky štátu na pokrytie jeho spotreby. Ako odmena za vzdanie sa likvidity je vlastníkom akceptovaný výnos. Štátne obligácie sú vyhľadávanými obligáciami pre spoločnosti, ako sú zaisťovne a poisťovne, ktoré ich zvyknú držať až do doby splatnosti. Tieto majú povinnosť ukladať svoje aktíva do zabezpečených cenných papierov. Je pravidlom, že štátne obligácie určujú výnosnosť aj ostatných obligácií emitovaných podnikmi na trhu. Štát sa stáva spoločnosťou, ktorej riziko platobnej neschopnosti je nižšie ako u súkromných spoločností. Obligácie emitované štátom sú menej rizikové ako obligácie ktoréhokoľvek podniku, ktorý vykonáva

svoju činnosť v danom štáte. Štátne obligácie Slovenskej republiky sa tešili donedávna veľkej pozornosti. Pri nízkom riziku ponúkali vysoký výnos. V sučasnosti sú už aproximované s výnosom eurozóny.

Cena obligácie

Cena obligácie odráža súčasnú hodnotu budúcich príjmov z obligácie. Tento vzťah sa dá vyjadriť aj rovnicou:

Cena obligácie = kupón/ (1 + výnos)^1 + kupón/ (1 + výnos)^2 + ... + kupón/ (1 + výnos)^n + istina/ (1 + výnos)^n

Kde **n** je počet období od emisie obligácie. Rast výnosu obligácie spôsobí pokles ceny obligácie a opačne. Treba si zároveň uvedomiť, že cena obligácie neodráža infláciu.

Alikvótny úrokový výnos predstavuje kompenzáciu kupónu v prospech predávajúceho. V čase medzi dvoma výplatami kupónov sa totiž predávajúci vzdáva kupónu v prospech kupujúceho.

Kupón je pravidelnou platbou, ktorú platí emitent obligácie majiteľovi obligácie.

Výnos do splatnosti vyjadruje úrokovú sadzbu, na základe ktorej sa súčasná hodnota cash-flow rovná cene.

Prečo sa cena obligácie vôbec mení? Treba si uvedomiť, že cena obligácie reflektuje najmä jeho výnos. Medzi výnosovou krivkou a cenou obligácie platí nepriama úmera. Ak výnosová krivka rastie, cena obligácie klesá a opačne. Toto pravidlo však platí, iba ak sú obligácie úročené fixný kupón.

Rád by som poukázal na závislosť medzi lehotou splatnosti a výnosom. V súčasnosti platí pravidlo, podľa ktorého je výnos priamo úmerný lehote splatnosti. Čím je teda lehota splatnosti obligácie dlhšia, tým investor očakáva vyšší výnos. Pre príklad uvádzam aukcie troch štátnych obligácií s rôznou lehotou splatnosti.

Číslo emisie	ISIN	Dátum emisie	Dátum splatnosti	Lehota splatnosti v rokoch	Menovitá hodnota [Sk]	Výnos v % p.a.	Perióda vypl. výnosov	Akceptované výnosy do splatnosti v % p.a.		
								min.	avg.	max.
ŠD202/A	SK4120004227	11.02.2004	11.02.2014	**10**	100,000	**4.90%**	ročne	5.009	5.145	5.180
ŠD203/A	SK4120004284	14.04.2004	14.04.2009	**5**	100,000	**4.80%**	ročne	4.890	4.898	4.900
ŠD204/A	SK4120004318	12.05.2004	12.05.2019	**15**	100,000	**5.30%**	ročne	5.160	5.273	5.300

Akceptovaný výnos súvisí najmä s absorbčnou schopnosťou trhu v momente aukcie.

Od 1.10.1996 má Burza cenných papierov v Bratislave, a.s.(BCPB) indikátor vývoja trhu obligácií - **Slovenský dlhopisový index - SDX**. Index je dvojzložkový - prvá zložka zahrňuje štátne obligácie a druhá obligácie spoločností

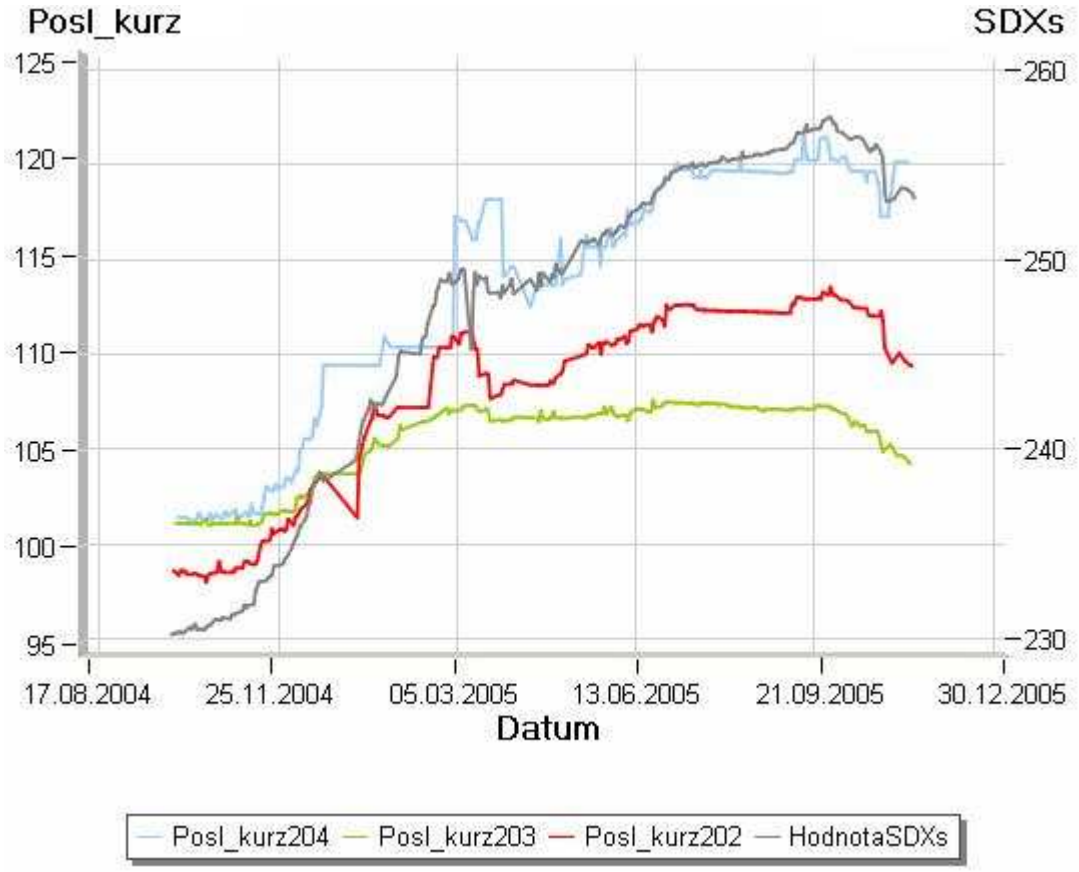
$$\mathbf{SDX} = K * \sum_{i=1}^{n} W_i * (P_i + AUV_i)$$

K = opravný koeficient
W_i = váha i-tej obligácie
P_i = indikatívna cena i-tej obligácie na BCPB
AUV_i = alikvótny úrokový výnos pre i-tu obligáciu v danom čase

Pre SDX bola po zvážení všetkých špecifík slovenského trhu obligácií ako najvhodnejšia zvolená kumulatívna formula, ktorá zahŕňa kapitálové zmeny (ceny), ako aj úrokové zmeny (alikvótny úrokový výnos a výplaty kupónov). Index je koncipovaný tak, aby predstavoval hodnotu 100 Sk investovaných na počiatku do portfólia indexu. Jeho hodnota rastie s pribúdajúcim akumulovaným úrokovým výnosom a rastie, alebo klesá s pohybom cien na trhu. Jednotlivé emisie v portfóliu indexu majú priradené váhy v pomere podľa celkovej veľkosti emisie.

SDX pre štátne obligácie zahŕňa v portfóliu indexu iba emisie štátnych obligácií. Jeho výpočet je zhodný s výpočtom indexu SDX.

V nasledujúcom grafe by som rád prezentoval vývoj kurzu štátnych obligácií ŠD 202, 203 a 204 za obdobie 1.októbra 2004 až 18. novembra 2005. Pre porovnanie je v grafe znázornený aj vývoj indexu SDX pre štátne obligácie. Rád by som upozornil na skutočnosť, že NBS znížila od 1.marca 2005 základnú úrokovú sadzbu zo 4% na 3%. Trh predpokladal tento vývoj, preto aj cena obligácií začala rasť. Znamená to, že trh akceptoval nižší výnos z obligácií. Najbližšie sa bude o základných úrokových sadzbách rokovať koncom decembra 2005. Trh v očakávaní zvýšenia týchto sadzieb reagoval znížením cien dlhopisov.



Korelácia výnosnosti obligácií

V mojom skúmaní vzájomnej korelácie výnosov štátnych obligácií ŠD202, 203 a 204 som vychádzal zo zhromaždených údajov za obdobie 10.januára 2005 až 18. novembra 2005. Za toto obdobie bolo dostupných 217 pozorovaní, ktoré slúžili ako báza dát pre zostavenie korelačnej matice. Do pozorovania som zahrnul aj vývoj indexu SDX pre štátne obligácie.

V nasledovnej tabuľke uvádzam základné štatistické parametre k jednotlivým zložkám štatistickeho súboru:

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
SDXs	217	251.35415	4.22121	54544	239.51000	257.15000	SDXs
Výnos202	217	0.03489	0.00288	7.57221	0.03054	0.04284	Výnos202
Výnos203	217	0.02940	0.00258	6.38015	0.02623	0.03800	Výnos203
Výnos204	217	0.03685	0.00271	7.99611	0.03185	0.04400	Výnos204

Po vykonaní základnej štatistiky som zostavil korelačnú maticu:

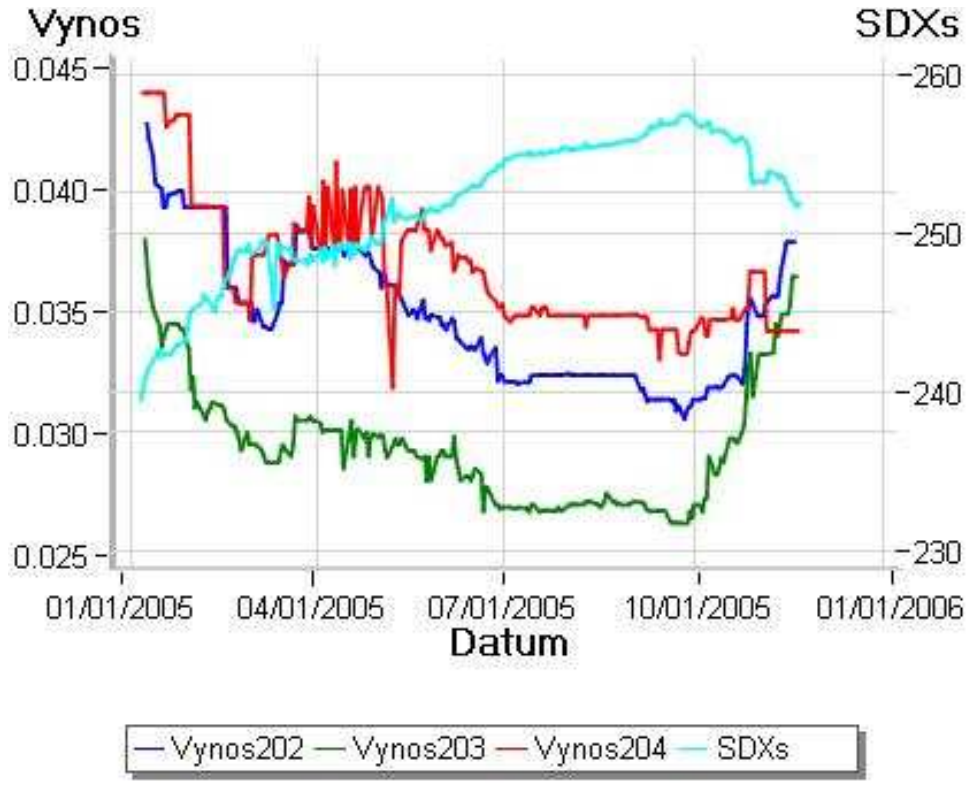
Pearson Correlation Coefficients, N = 217

	SDXs	Výnos202	Výnos203	Výnos204
SDXs	1.00000	-0.93907	-0.71923	-0.89596
Výnos202	-0.93907	1.00000	0.81753	0.82687
Výnos203	-0.71923	0.81753	1.00000	0.60864
Výnos204	-0.89596	0.82687	0.60864	1.00000

Z matice vyplývajú vzájomné vzťahy medzi parametrami:

a) Výnos_ŠD202 / Výnos_ŠD203 = 0.81753 => pozitívna korelácia
b) Výnos_ŠD202 / Výnos_ŠD204 = 0.82687 => pozitívna korelácia
c) Výnos_ŠD203 / Výnos_ŠD204 = 0.60864 => pozitívna korelácia
d) SDX_s / Výnos_ŠD202 = -0.93907 => negatívna korelácia
e) SDX_s / Výnos_ŠD203 = -0.71923 => negatívna korelácia
f) SDX_s / Výnos_ŠD204 = -0.89596 => negatívna korelácia

Pre názornosť výsledkov uvádzam graf, ktorý zachytáva vývoj výnosov sledovaných štátnych obligácií a indexu SDX pre štátne obligácie v časovom horizonte 10.januára 2005 až 18. novembra 2005:



Vývoj hodnôt potvrdzuje výsledky korelačnej matice a tým aj hypotézu o vzájomnej pozitívnej závislosti výnosov sledovaných obligácií.

Záver

Z uskutočnených pozorovaní a korelačnej analýzy vyplýva, že existuje pozitívna korelácia medzi výnosmi štátnych obligácií ŠD202, 203 a 204. Znamená to, že výnosy ležia na priamke s kladnou smernicou. Pre portfóliových manažérov znamená tento údaj potvrdenie skutočnosti, že existuje kladná závisloť medzi výnosmi uvedených štátnych obligácií. Rozdiel medzi jednotlivými obligáciami je v lehote splatnosti a s tým súvisiacim výnosom.

Zároveň som dokázal, že medzi výnosmi a indexom SDX pre štátne obligácie je negatívna korelácia, čo potvrdzuje pravidlo nepriameho vzťahu ceny obligácie a výnosu do splatnosti. Zníženie výnosnosti pôsobí na zvýšenie ceny obligácie a opačne.

Použitá literatúra:

1) Chovancová B., Jankovská A., Kotlebová J., Šturc B.: *Finančný trh, Eurounion 2002*
2) INVESTOR, november 2005, ročník VI.
3) www.ardal.sk
4) www.nbs.sk
5) www.etrend.sk
6) www.tatrabanka.sk
7) Bloomberg
8) EBOS – systém obchodovania Burzy cenných papierov v Bratislave, a.s.

Vypracoval: Branislav Rajčáni, FHI EUBA, IT4, 4.ročník

VÝVOJ POTRATOVOSTI NA SLOVENSKU PODĽA DOSIAHNUTÉHO VEKU ŽENY

Ivana Rajecová

Úvod

Dôležitým ukazovateľom úrovne potratovosti žien (samovoľnej, umelej a celkovej potratovosti) je dosiahnutý vek.

S vekom ženy sa výrazne mení podiel potratov na celkovom počte ukončených tehotenstiev. Samozrejme v rôznych vekových kategóriách zaznamenávame rôznu intenzitu potratovosti.

Na Slovensku je evidencia potratov triedená podľa dokončeného veku ženy a druhu potratu, čo sme využili na detailnejšiu analýzu špecifickej potratovosti. Pritom sme špecifikovali podľa toho, či ide o samovoľné potraty alebo umelé potraty.

Vývoj miery potratovosti podľa veku

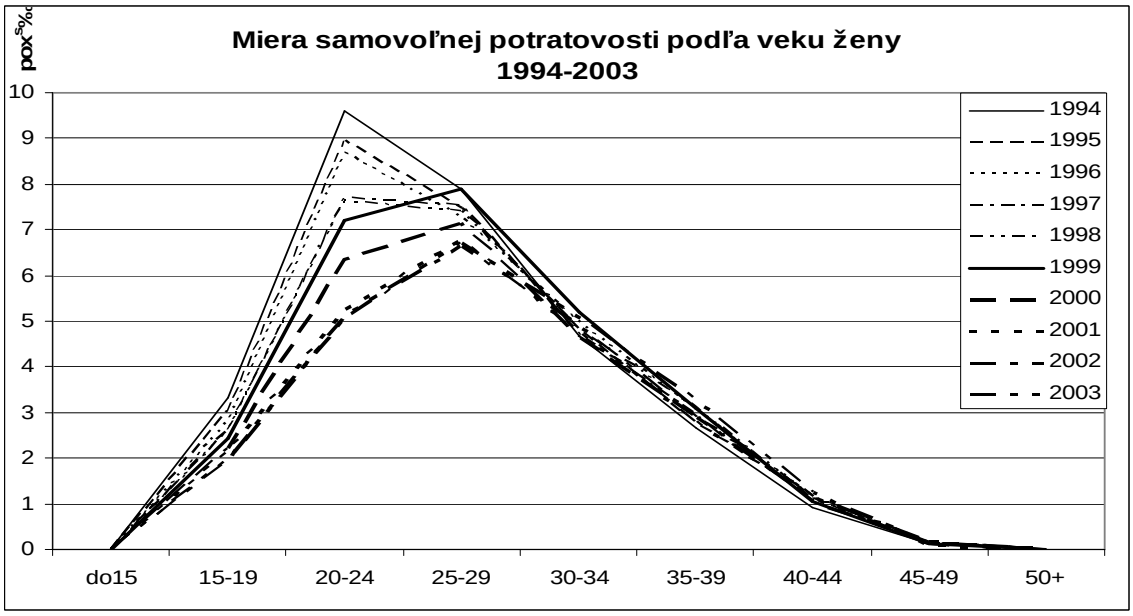
Počas sledovaného vývojového obdobia od roku 1994 až do roku 2003, všeobecne sledujeme pozitívny klesajúci vývojový trend špecifickej potratovosti.

Od roku 1994 zaznamenávame pokles potratovosti vo všetkých vekových kategóriách.

Na začiatku sledovaného obdobia pozorujeme úzky vrchol spontánnej potratovosti. Tento vrchol bol zaznamenaný u žien vo vekovej kategórii 20-24 ročných. Miera špecifickej spontánnej potratovosti (počet potratov v danom veku x za rok/na 1000 x - ročných žien stredného stavu) v tejto vekovej kategórii sa pohybovala na úrovni 9,6 spontánnych potratov na 1000 žien fertilného veku.

Ako je vidieť z grafu č.1 v roku 1995 a 1996 sledujeme pokles miery spontánnej potratovosti, avšak maximum je stále vo vekovej kategórii 20-24 ročných žien.

Graf č.1:



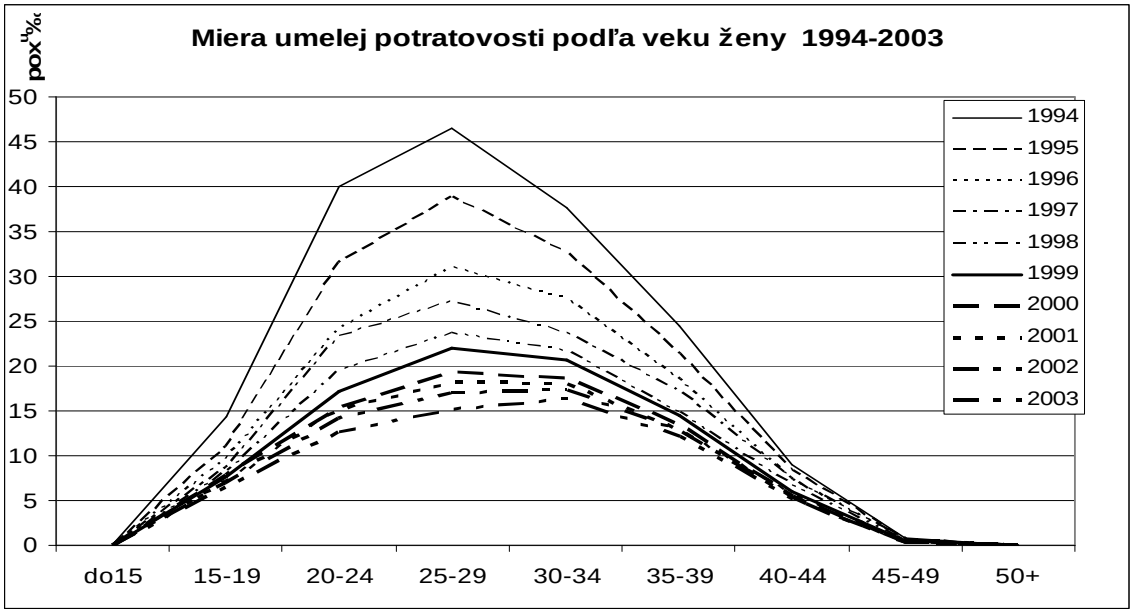
Toto sa mení v roku 1997, kedy po prvýkrát sledujeme 2 viditeľné maximá spontánnej potratovosti vo vekovej kategórii 20-24 ročných a 25-29 ročných. Špecifická miera spontánnej potratovosti bola v tom čase na úrovni 7,6 promile. Nárast spontánnej

potratovosti pozorujeme v roku 1999 (7,9 ‰), kde sledujeme presun maxima spontánnej potratovosti do vyššej vekovej kategórie 25-29 ročných žien.
Toto maximum je zachované aj v súčasnosti. V roku 2003 sa pohybovala miera spontánnej potratovosti vo vekovej kategórii 25-29 ročných na úrovni 6,7 ‰.

Tak ako pri spontánnej potratovosti aj úroveň špecifickej miery umelej potratovosti sa znížila vo všetkých vekových kategóriách žien.
Pozitívne možno hodnotiť najmä výrazné zníženie počtu interrupcií u žien v najmladšej vekovej kategórií 15-19 ročných.

Na začiatkom sledovaného obdobia vykazovali najvyššiu mieru umelej potratovosti ženy vo veku 25 - 29 rokov. V roku 1994 predstavovala hodnotu umelej potratovosti v tejto vekovej kategórii 55,7 ‰. Druhú najpočetnejšiu kategóriu v roku 1994, tvorila kategória 20-24 ročných s 51 interrupciami na 1000 žien.
Podobný priebeh majú aj krivky v roku 1995, 1996 a 1997, ktoré sa od seba líšia klesajúcou špecifickou mierou umelej potratovosti. V roku 1997 sa znížila hodnota špecifickej miery umelej potratovosti vo vekovej kategórii 25-29 ročných na hodnotu 35,7 ‰, vo vekovej kategórii 20-24 to bolo 31,4 ‰ a v poslednej kategórii 30-34 ročných bola hodnota miery umelej potratovosti 29,1 ‰.
Dlhodobo najviac interrupcií je koncentrovaných práve do vekovej skupiny 25-29 ročných žien.
V roku 1998 sa zmenilo poradie druhej a tretej najpočetnejšej vekovej kategórie. O tohto roku po vekovej kategórii 25-29 ročných, nasledujú vekové kategórie žien 30-34 ročných a až po nej nasleduje kategória 20-24 ročných. V súčasnosti až 68% všetkých interrupcií vykazujú ženy práve týchto troch vekových skupinách (2004, D.Jurčová).

Graf č.2:



Najnižšiu úroveň umelej potratovosti si udržiavajú okrajové vekové kategórie žien. Najmladšie ženy do 19 rokov, ktoré práve vstupujú do reprodukčného obdobia a ženy nad 40 rokov, ktoré sú už takmer na konci svojho reprodukčného obdobia.
Najväčší pokles umelej potratovosti od začiatku sledovaného obdobia až po súčasnosť bol zaznamenaný práve u žien 25 – 29 ročných, kde bol pokles viac ako trojnásobný.

Je to určite do veľkej miery spôsobené lepšími preventívnymi opatreniami zo strany týchto žien, ako i menším počtom celkovo realizovaných tehotenstiev (2003,

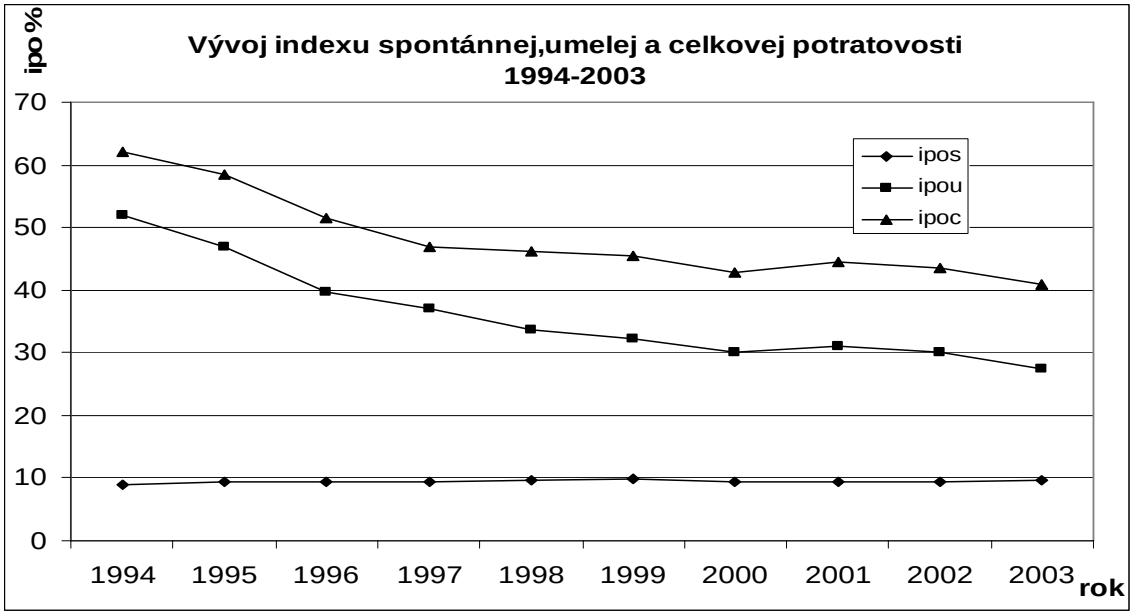
B.Vaňo). S rastúcim vekom žien sa tempo i úroveň poklesu potratovosti znižuje, ale klesajúce tendencie pretrvávajú vo všetkých vekových kategóriách. Pokles u najmladšej vekovej kategórie žien do 19 rokov treba hodnotiť veľmi pozitívne, najmä preto, že naznačuje zodpovednejšie správanie mladej generácie, ktorá sa spolieha skôr na antikoncepciu ako na interrupciu (2003, B.Vaňo).

Do roku 2003 sa úroveň umelej potratovosti znížila vo všetkých vekových kategóriách žien. Zároveň vidíme mierne posúvanie interrupcií do vyššieho veku, čo súvisí hlavne so zvyšovaním priemerného veku žien pri sobáši a pôrode. Zatiaľ čo v roku 1995 bol priemerný vek ženy pri interrupcii 28,22 do roku 2003 sa zvýšil na hodnotu 29,1.V roku 2003 už neexistuje žiadne výrazné maximum úrovne umelej potratovosti podľa veku, krivka umelej potratovosti za tento rok je omnoho vyrovnanejšia a oblejšia ako tomu bolo v roku 1994.

Vývoj indexu potratovosti podľa veku

Samovoľnú potratovosť ako biologický jav najlepšie charakterizuje index samovoľnej potratovosti (počet samovoľných potratov/ na 100 živo narodených). Tento index je v SR relatívne stály, kolíše okolo 10 % v celom sledovanom období.

Graf č.3:

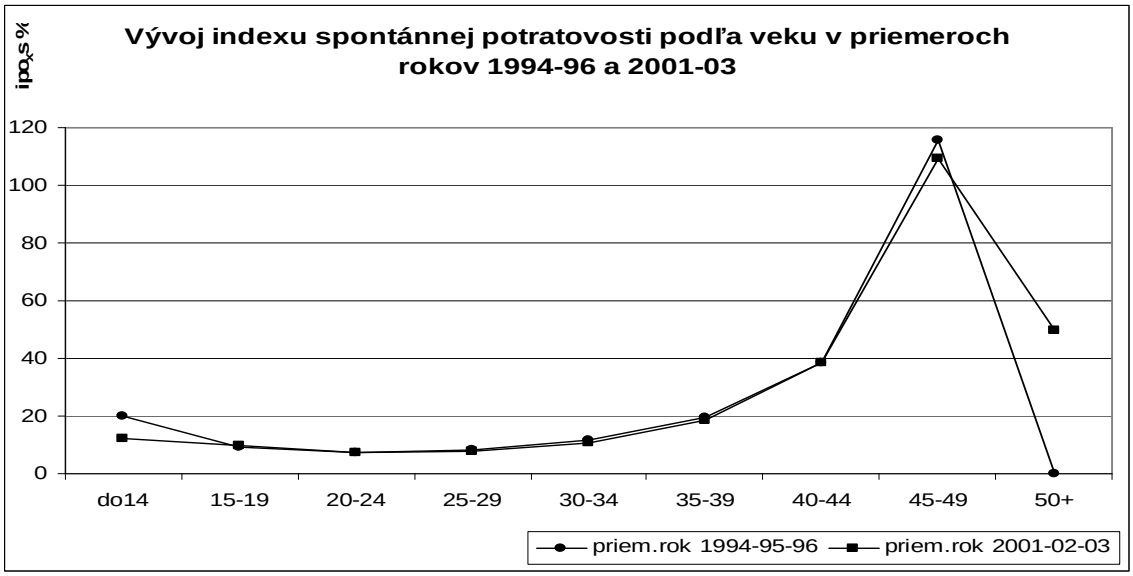


Ako je vidieť z grafu č.4, sklon k samovoľnej potratovosti je najmenší u žien vo veku 18 -29 rokov. V priemere sa pohybuje úroveň indexu spontánnej potratovosti v týchto vekových kategóriách počas celého skúmaného obdobia na úrovni 7,5%.
Na vývoji samovoľnej potratovosti sa prejavuje jej silná biologická podmienenosť a teda index špecifickej spontánnej potratovosti rastie s pribúdajúcim vekom ženy. So zvyšujúcim sa vekom ženy sa znásobujú možné zdravotné riziká a iné komplikácie.

Počas sledovaného obdobia pozorujeme aj v prípade samovoľnej potratovosti jej mierny pokles. Ten sa však sústredil len na najmladšie vekové skupiny 1 8 – 27 rokov a bol prakticky spôsobený len poklesom počtu tehotenstiev. Vo veku, kde počet tehotenstiev neklesol, si samovoľná potratovosť zachovala z minulosti svoj trend a prakticky aj úroveň . Vo veku 45 rokov až 60% prirodzene ukončených tehotenstiev sa končí samovoľným potratom.

Vo vekovej kategórii 45-49 ročných prevyšujú spontánne potraty nad pôrodmi. Je potrebné mať samozrejme na zreteli nízke absolútne počty tehotenstiev v tomto veku (na Slovensku bolo v roku 1999 evidovaných 1 40 tehotenstiev vo veku nad 45 rokov).

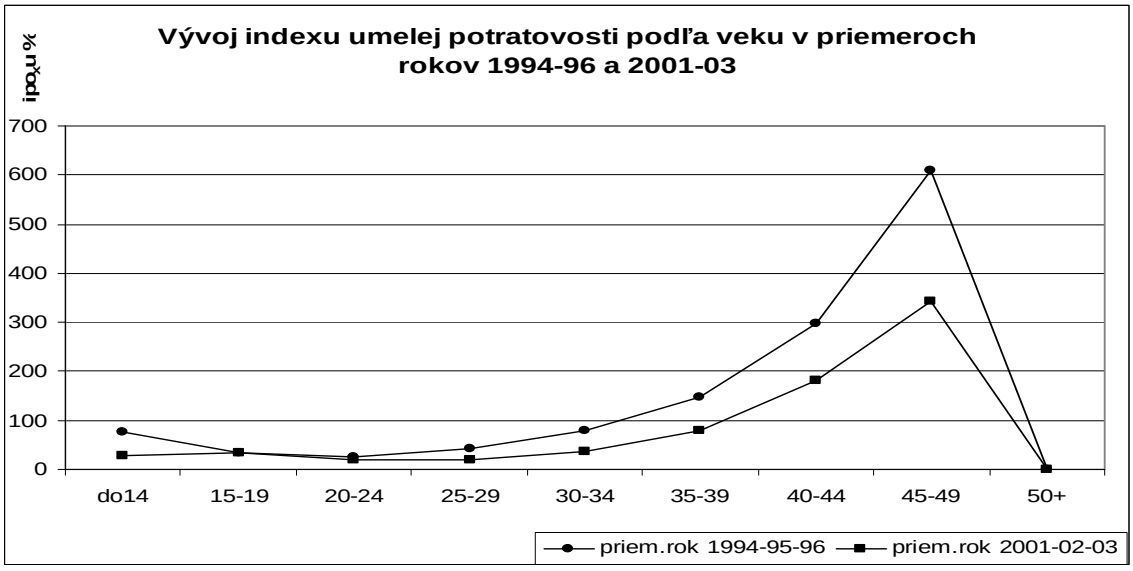
Graf č.4:



V sledovanom období rokov 1994 až 2003 pozorujeme výrazný pokles indexu umelej potratovosti. Zatiaľ čo v roku 1994 sa pohyboval index umelej potratovosti na úrovni 51,9 interrupcii na 100 živo narodenýchm, v roku 2003 to bolo skoro o polovicu menej 27,4 interrupcii100 živo narodených.

Ak sa pozrieme na index umelej potratovosti podľa veku zistíme, že úroveň umelej potratovosti klesla vo všetkých vekových kategóriách. Zvýšený nárast indexu umelej potratovosti v najnižšej vekovej kategórii je spôsobený nízkym počtom tehotenstiev v tejto vekovej kategórii. Ide o veľmi mladé dievčatá pre ktoré je jediným východiskom interrupcia. Preto pokles interrupcii, ktorý pozorujeme v posledných rokoch v tejto vekovej kategórii považovať za výrazne pozitívny jav.

Graf č.5:



Najnižší index umelej potratovosti zaznamenávame vo vekových kategóriách 20-24 ročných v celom skúmanom období (index umelej potratovosti sa tu pohybuje na úrovni 21%).
Mierny nárast indexu umelej potratovosti zaznamenávame vo vekovej kategórii 25-29 ročných žien. V roku 1994 bol index umelej potratovosti v tejto vekovej kategórii 48,9%, zatiaľ čo v roku 2003 to bolo len 17,8%. Až do veku 30 rokov výrazne prevažujú medzi ukončenými tehotenstvami pôrody.
Výraznejší nie veľmi povzbudivý nárast sledujeme vo vekovej kategórii 30-34 ročných.. Po tejto vekovej kategórii začína výrazne narastať podiel interrupcií na úkor pôrodov. Výrazný zlom nastáva okolo 37 rokov, od ktorého sa interrupcie stávajú najčastejším spôsobom ukončenia tehotenstva. Na konci reprodukčného obdobia výrazne prevyšujú interrupcie nad pôrodmi. Ako sme už spomínali ide o nízke absolútne počty tehotenstiev v týchto vekových kategóriách.

Záver

Počas celého skúmaného obdobia od roku 1994 až 2003, všeobecne sledujeme pozitívny klesajúci vývojový trend špecifickej potratovosti.
Pokles potratovosti zaznamenávame vo všetkých vekových kategóriách žien, čo sme preukázali pomocou miery potratovosti a indexu potratovosti podľa veku.
Najdôležitejším, výrazne pozitívnym javom, ktorý pozorujeme v posledných rokoch je pokles umelej potratovosti vo všetkých vekových kategóriách, ale najmä pokles interrupcii v najmladších vekových kategóriách do 15 rokov a 15-19 ročných žien.

Literatúra

- D.Jurčová, M.Lukáčová, J.Mészároš, V.Pilinská, Demografická charakteristika obvodov Slovenskej republiky 1996-2003, VDC, 2004,BA
- B.Vaňo a kol., Populačný vývoj Slovenska, VDC, 2003, BA
- Pavlík, Z., Rychtaříková, J., Šubrtová, A.: Základy demografie. 1.vyd. Praha, Academia 1986
- Potraty v SR 1994, Ústav zdravotníckych informácii a štatistiky, 1995, Bratislava
- Potraty v SR 1995, Ústav zdravotníckych informácii a štatistiky, 1996, Bratislava
- Potraty v SR 1996, Ústav zdravotníckych informácii a štatistiky, 1997, Bratislava
- Potraty v SR 1997, Ústav zdravotníckych informácii a štatistiky, 1998, Bratislava
- Potraty v SR 1998, Ústav zdravotníckych informácii a štatistiky, 1999, Bratislava
- Potraty v SR 1999, Ústav zdravotníckych informácii a štatistiky, 2000, Bratislava
- Potraty v SR 2000, Ústav zdravotníckych informácii a štatistiky, 2001, Bratislava
- Potraty v SR 2001, Ústav zdravotníckych informácii a štatistiky, 2002, Bratislava
- Potraty v SR 2002, Ústav zdravotníckych informácii a štatistiky, 2003, Bratislava
- Potraty v SR 2003, Ústav zdravotníckych informácii a štatistiky, 2004, Bratislava

Ivana Rajecová, UK Prírodovedecká fakulta, 5. ročník

OKEČ a jeho vplyv na výsledok hospodárenia podniku

Pavol Skalák

Úvod

Každé ekonomické odvetvie, v ktorom podnik pôsobí má svoje špecifické črty (rozdiel napr. v konkurencii, dodávateľoch, dopyte). Odvetvie by malo ovplyvňovať jeho ekonomické postavenie, ktoré sa môže prejaviť napríklad vo výsledku hospodárenia daného podniku. Tento vplyv môžeme analyzovať pomocou faktora *OKEČ* a premennej *výsledok hospodárenia*.

Podkladom pre analýzu boli údaje, ktoré poskytla firma *Infin s.r.o*. Ide o účtovné závierky podnikateľských subjektov na Slovensku za rok 2004 z ktorých bol použitý práve výsledok hospodárenia. Analýza bola urobená v štatistickom systéme *SAS* (verzia 9), ktorý patrí k svetovej špičke. Z pôvodného štatistického analytického systému sa stal špičkovým systémom metód a postupov na podporu rozhodovania. Je zložený z viacerých modulov, v štatistike sa využívajú najmä moduly BASE, STAT, ETS a GRAPH. SAS ponúka aj širokú paletu oceňovaných a široko využívaných riešení pre manažment údajov a iných analytických riešení.

Analýza vplyvu OKEČ na výsledok hospodárenia

Výsledok hospodárenia podniku je rozdielom medzi všetkými výnosmi a nákladmi podniku. Keď sú výnosy vyššie ako náklady, ide o zisk, ak je to opačne, je výsledkom strata. Zisk ako finančný výsledok podnikovej činnosti je významným kvalitatívnym ukazovateľom. Charakterizuje efektívnosť používanie podnikového kapitálu, je zdrojom dôchodkov vlastníkov a dôležitým zdrojom financovania rozvoja podniku.
Treba si uvedomiť, že výšku zisku za jednotlivé obdobia možno v praxi ovplyvniť rôznymi zásahmi, ktoré majú viac menej účtovnícky charakter (napr. spôsob časového rozlíšenia nákladov, spôsob oceňovania zásob, voľba spôsobu odpisovania investičného majetku). Pri riadení podniku sa zisk využíva nielen ako zdroj financovania, ale aj ako jeden z najdôležitejších syntetických ukazovateľov úspešnosti a efektívnosti podnikania (najmä vo vzťahu k vloženému kapitálu) ako aj nástroj hmotnej zainteresovanosti (tak majiteľov, manažérov, ako aj pracovníkov). Výška dosiahnutého výsledku hospodárenia (zisku alebo straty) závisí od niekoľkých základných činiteľov:

- Objem realizovanej produkcie
- Štruktúra realizovanej produkcie
- Cena jednotky realizovanej produkcie
- Náklady na jednotku realizovanej produkcie

OKEČ (Odvetvová Charakteristika ekonomických činností) charakterizuje činnosť firmy. Hlavným účelom odvetvovej klasifikácie je poskytnúť hierarchické triedenie ekonomických činností, ktoré sa môžu využiť na rozčleňovanie informácií podľa činností na rôznych úrovniach agregácií v analytických prácach v štatistike a v iných oblastiach. Na analýzu sa použil dvojmiestny OKEČ, ktorý sa označuje ako oddiel (napr. priemyselná výroba, ťažba nerastných surovín, stavebníctvo a iné).

Zaujíma nás, či faktor OKEČ má vplyv na výšku výsledku hospodárenia. Na zistenie tejto závislosti je najvhodnejšie použiť metódu ANOVA (ANalysis of VAriance) - analýza rozptylu. Je to metóda na porovnávanie priemerov niekoľkých základných súborov a jej zmyslom je určiť, či sú alebo nie sú rozdiely medzi priemermi týchto súborov.

Úlohou je špecifikovať, či úrovne faktora OKEČ štatisticky významne ovplyvňujú variabilitu premennej výsledok hospodárenia. Ak by sa priemerné výšky VH rovnali, potom faktor OKEČ nemá vplyv na priemernú výšku VH. Skôr ale môžeme očakávať práve opačnú situáciu, t.j. priemerné hodnoty budú závisieť od úrovní OKEČ. Pri analýze rozptylu overujeme hypotézu:

H_0: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_r$

H_1: aspoň dva priemery sa nerovnajú

Výsledkom analýzy rozptylu je tabuľka analýzy rozptylu v ktorej sú usporiadané podklady na rozhodnutie o výsledku testu.

Celý súbor obsahuje 44 672 subjektov, ktoré sú rozložené do 57 oddielov OKEČ. Veľkosti jednotlivých výberových súborov nie sú rovnaké, to znamená že ide o nevyvážený model. Pomocou tabuľky č.1, môžeme teda rozhodnúť o prijatí alebo zamietnutí nulovej hypotézy.

Tab č.1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	56	3.6068079E14	6.4407283E12	1918.46	<.0001
Error	44615	1.4978319E14	3357238382.8		
Corrected Total	44671	5.1046398E14			

R-Square	Coeff Var	Root MSE	VH Mean
0.706574	6171.562	57941.68	938.8495

Posledný stĺpec tabuľky obsahuje p-hodnotu (Pr > F), ktorá predstavuje najnižšiu hladinu významnosti pre zamietnutie nulovej hypotézy. Pre prijatie nulovej hypotézy by sme museli zvoliť hladinu významnosti α nižšiu ako táto p-hodnota. V tabuľke je p-hodnota menšia ako 0.0001, to znamená, že na hladine významnosti α = 0.05 nulovú hypotézu zamietame a prijímame predpoklad, že priemerné hodnoty VH závisia od úrovní faktora OKEČ. Pre takto malú p-hodnotu by sme mohli H_0 zamietnuť na akejkoľvek prijateľnej hladine významnosti.
Hodnota R-Square nám vyjadruje, aký podiel z celkovej variability predstavuje modelom (faktorom) vysvetlená variabilita. Uvedená hodnota 0.706574 hovorí, že vysvetlených je 70.65% celkovej variability a 29.35 je náhodná zložka.

Pre dokreslenie obrazu o súbore si môžeme pozrieť sumárne štatistiky v tabuľke č.2.

Tab. č.2

Okec	N	Minimum	Maximum	Mean	Median	Std Dev
	44672	*-18446578*	*9693486*	*938.85*	*2.0*	*106898.03*
00	631	-23704	8187	-70.838	-13.0	1436.952
01	1318	-174374	30894	-91.005	82.0	7147.563
02	160	-2275	179898	1478.788	41.5	14252.181
05	11	-1167	5985	808.909	148.0	1909.764
10	5	-6784	712	-1140.200	-2.0	3171.056

11	2	-2758	-30	-1394.000	-1394.0	1928.987
13	2	982	2082	1532.000	1532.0	777.817
14	76	-85603	32730	-974.934	24.0	14003.372
15	685	-119617	385091	1072.534	0.0	25503.027
16	2	-21164	319053	148944.500	148944.5	240569.748
17	157	-239595	78591	-980.159	0.0	20910.413
18	251	-103902	79196	-293.817	0.0	8958.794
19	118	-49155	179573	2098.593	30.5	18869.407
20	614	-111841	171040	128.463	0.0	10646.703
21	66	-5972	648643	18208.485	177.5	81614.407
22	497	-66597	88608	895.364	22.0	7844.300
23	5	-28416	9693486	1939605.000	-2.0	4334625.116
24	143	-242202	446816	5518.657	11.0	58063.710
25	350	-167568	208515	1276.374	1.5	20753.997
26	296	-72085	720708	9666.584	1.5	55375.759
27	54	-212976	1366042	40753.722	522.0	230967.019
28	1003	-406089	90880	486.393	52.0	14660.176
29	555	-258202	831532	3021.175	107.0	42739.078
30	47	-1246	88676	3494.319	82.0	13165.509
31	336	-51447	193074	1947.467	98.0	14642.253
32	119	-24039	118243	3260.227	70.0	14807.531
33	185	-6367	50121	1549.897	27.0	5792.130
34	81	-317972	5301620	77870.296	25.0	592556.204
35	41	-25375	125389	3037.805	23.0	21302.602
36	328	-442329	252159	-814.421	0.0	30862.319
37	93	-14271	18465	578.108	4.0	3560.437
40	119	-86948	2626955	58106.924	857.0	319040.321
41	23	-583320	187128	-13236.174	44.0	130369.151
45	3274	-44143	295294	695.673	12.0	7492.994
50	973	-172427	135197	918.121	6.0	10539.376
51	9842	-574237	1013168	783.017	0.0	14684.949
52	6798	-254962	236855	105.847	0.0	7531.647
55	1299	-324429	120794	-374.092	-15.0	9970.999
60	823	-146028	887148	1613.770	0.0	35825.643
61	10	-65	31023	6370.000	296.5	12307.643
62	9	-6635	25934	2237.111	0.0	9194.640
63	734	-16361	265228	1204.823	8.0	12246.414
64	89	-28416	4141143	75043.652	33.0	479231.437
65	142	-7732	385769	9422.697	0.0	44031.159
66	5	-745	92024	34859.400	24930.0	39972.992
67	152	-81899	112344	1224.224	21.5	16032.101
70	2021	-139242	112036	-106.671	0.0	7972.512
71	418	-13354	201873	1386.677	0.0	11722.794
72	1144	-17259	173856	1327.302	28.0	8656.315
73	102	-24721	56917	967.794	3.0	7268.054
74	6871	-1172676	188818	93.995	5.0	15503.690
75	1	-18446578	-18446578	-18446578.00	-18446578.0	.
80	232	-3684	5092	70.086	1.5	771.450
85	420	-57687	50174	99.610	8.0	5068.513
90	160	-18003	22561	1190.144	141.5	3866.810
92	568	-133923	448341	944.072	0.0	21913.036
93	212	-46127	5482	-339.014	-6.5	3468.055

* hodnoty sú v tis. SK

V tabuľke vidíme, že rozloženie početnosti premennej výsledok hospodárenia je v jednotlivých OKEČoch veľmi rozdielne. Od jedného (OKEČ 75) až po 9842 pozorovaní (OKEČ 51). Ďalej môžeme konštatovať, že hodnota mediánu je výrazne nižšia ako hodnota priemeru. To znamená, že počet podnikov s nižším výsledkom hospodárenia je výrazne väčší ako počet podnikov s vyšším výsledkom hospodárenia. Minimálnu hodnotu premennej dosiahol OKEČ 75 a maximálnu hodnotu OKEČ 23.

Záver

Podnik s najvyššou stratou sa nachádza v odvetví *„Verejná správa a obrana; povinné sociálne zabezpečenie“*. Najmenší počet pozorovaní sa nachádzal práve v tomto OKEČ. Najvyšší zisk sa dosiahol v odvetví *„Výroba koksu, rafinovaných ropných produktov a jadrového paliva“*.

Existenciu vplyvu medzi činnosťou podniku a jeho hospodárením sa nám podarilo dokázať analýzou rozptylu. Tá ukázala vzájomnú závislosť OKEČ a výsledku hospodárenia.

Literatúra

1) Chajdiak J., Komorník J., Komorníková M.: Štatistické metódy, Bratislava, Statis, 1999
2) Chajdiak J., Rublíková E., Gudába M.: Štatistické metódy v praxi, Bratislava, Statis, 1994
3) Chajdiak J.: Štatistika jednoducho, Bratislava, Statis, 2003
4) Vlachynský K. a kol.: Podnikové financie, Bratislava, Súvaha, 2002
5) Chajdiak J. a kol.: Príručka pre užívateľov systému SAS, Bratislava, Statis, 1994

Pavol Skalák
4. ročník, Informačné technológie, FHI, EU Bratislava

Modelovanie funkcie konečnej spotreby domácností v SR

František Štulajter

Abstrakt

The case study focuses on macroeconomics dimensions as consumption of households, gross domestic product, inflation and interest rate. The purpose is to estimate a regression analysis of total household´s consuption during the period from year 1995 till the midle of 2005 and to summarize forecasting possibilities.

Úvod

Práca vychádza zo základných ekonomických predpokladov, že spotreba domácností rastie s rastúcimi mzdami ako aj s rastúcim importovaným tovarom a službami a tiež s rastom výstupu ekonomiky, ale klesá vplyvom rastúcej inflácie, nezamestnanosti a bankovej úrokovej miery.

Za cieľ práce sme si stanovili vytvorenie modelu pomocou regresnej analýzy, ktorý by poskytoval možnosti predikcie budúcej konečnej spotreby domácností.

Práca si nekladie za cieľ vyčerpávajúco použiť všetky dostupné štatistické metódy a ani popis ich výpočtových algoritmov.

Práca bola vytvorená použitím študentskej verzie softvérového programu EViews 4.1.

Vlastná práca je rozdelená do dvoch častí:

- vytvoriť klasický lineárny model pomocou metódy najmenších štvorcov OLS s ohľadom na splnenie predpokladov klasického lineárneho modelu
- otestovať predikčnú schopnosť modelu

Tabuľka číslo 1 poskytuje popis vstupných dát ich zdroj a periodicitu

Tab.č.1

Ozn.	Popis	M.j.	Zdroj	Periodicita dát
SP	Konečná spotreba domácnosti	mil.SKK	www.nbs.sk www.slovstat.sk	Kvartálna
IM	Import	mil.SKK		Kvartálna
INF	Inflácia	% p.a.		Kvartálna
MZ	Priemerná mesačná mzda zamestnanca	SKK		Kvartálna
U	Nezamestnanosť	%		Kvartálna
I	Úroková miera na netermínovaných vkladoch	% p.a.		Kvartálna
GDP	Hrubý domáci produkt v b.c. roku 1995 (ESNÚ95)	mil.SKK		Kvartálna

Dáta boli rozdelené do dvoch skupín, do prvej skupiny bolo zaradené obdobie od prvého kvartálu 1995 do druhého kvartálu 2002. Na tejto skupine bol vytvorený odhadný model a na druhej skupine, obdobie od tretieho kvartálu 2002 do polovice roku 2005, uskutočnená a zhodnotená predikčná schopnosť modelu.

Prvý model má špecifikáciu ako funkčná závislosť SP≈ f(c,IM,INF,MZ,U,I,GDP), kde c je lokujúca premenná. Tento model má dve obmedzenia a síce to, že SP je pravdepodobne nestacionárna veličina a medzi regresormi sa vyskytujú vysoké lineárne závislosti, čo je potvrdené korelačnou maticou. Rovnako skúmané makroekonomické veličiny periodicky kolíšu počas kalendárneho roka.

Problém nestacionarity SP sme sa rozhodli riešiť zmenou špecifikácie modelu pomocou diferencií prvého rádu d(SP)= f(c,d(IM),d(INF),d(MZ),d(U),d(I),d(GDP)). Stacionaritu

exogénnych a endogénnych premenných sme skúmali pomocou *rozšíreného Dickey-Fuller testu(ADF).*

ADF test predpokladá nulovú hypotézu H_o: $\rho = 1$, ktorá hovorí o nestacionarite časového radu[1] a je testovaná proti jednostrannej alternatíve H_1: $\rho < 1$. Dickey a Fuller (1979) prezentujú, že H_o nemá štandardné rozdelenie funkcie pravdepodobnosti. Boli použité tabelované hodnoty Critical Value (MacKinnon, 1991). H_o zamietame na danej hladine významnosti ak |ADF Test Statistic| > Critical Value.

ADF test potvrdil hypotézu o nestacionarite spotreby obyvateľstva v sledovanom období a taktiež potvrdil stacionaritu prvej diferencie spotreby d(SP), ktorá je na všetkých bežných hladinách významnosti stacionárna.

Problém multikolinearity sme riešili pomocou korelačnej matice regresorov, na základe ktorej sme zo súboru premenných s vysokou priamou lineárnou závislosťou ponechali len jednu.

Na základe Akaikovho informačného kritéria (AIC) a Schwarzovho kritéria (SC) sme vybrali tú premennú, ktorej ponechanie malo najlepší výsledok pre regresnú analýzu.

Na riešenie problému sezónnosti, ktorá sa pravidelne objavuje v sledovanom modely, sme sa rozhodli použiť sezónne autoregresné procesy a sezónne procesy kĺzavých priemerov. Pod sezónnosťou sa rozumie periodické kolísanie, ktoré má systematický charakter a väčšinou sa opakuje behom jedného kalendárneho roka. V prípade ekonomických časových radoch sa predpokladá, že sezónna zložka má stochastický charakter.

K odstráneniu problému sezónnej korelovanosti rezíduí došlo použitím procesov SARIMA v odhadovanom modely. Následne sme použili diferenciu vyššieho rádu v prípade vysvetľujúcej premennej IM. Jednalo sa o štvrtú diferenciu definovanú ako nárast resp. pokles objemu dovezených tovarov a služieb počas štyroch štvrťrokov.

V ďalšom kroku došlo k testovaniu špecifikácie modelu *Ramsey's RESET Test.* RESET test, čiže *Regression Specification Error Test Ramsey (1969),* je všeobecne používaný test na zisťovanie špecifikačných chýb modelov, ktoré vznikli vynechaním vysvetľujúcej premennej alebo chybným analytickým tvarom modelu a rovnako upozorňuje na možnú koreláciu medzi vysvetľujúcimi premennými a rezíduami. Nulová hypotéza je H_o: $\xi \sim N(0,\delta^2 I)$. Ako špecifikačný faktor sme použili "1", teda vloženie premennej $\hat{y}^2$ do odhadovanej regresie.

Tab.č.2

F-statistic	0.973405	Probability	0.338518
Log likelihood ratio	1.299298	Probability	0.254342

Ramsey's RESET Test nepreukázal špecifikačné chyby použitého modelu. Na testovanie stability odhadovaného modelu neboli použité testy odhadov rekurzívnych rezíduí z dôvodu použitia autoregresných procesov SARIMA.

Jednou z podmienok uskutočnenia inferencie a použitia OLS odhadov je tiež normálne rozdelenie náhodných chýb $\xi \sim N(0,\delta^2 I)$. Na tieto účely bol použitý test *Jarque-Bera*. Test porovnáva rozdiely medzi špicatosťou a šikmosťou rezíduí modelu a normálneho rozdelenia. Za platnosti nulovej hypotézy o normalite rezíduí má *Jarque-Bera štatistika* chí-kvadrát rozdelenie s dvoma stupňami voľnosti.

Tab.č.3

Jarque-Bera	1.543264
Probability	0.462258

Na základe výsledkov *Jarque-Bera štatistiky* nemôžeme zamietnuť nulovú hypotézu.

[1] Ak je platí, že : $|\rho| > 1$, jedná sa o explozívny časový rad, ktorý nemá ekonomickú vypovedaciu schopnosť

Porušenie predpokladu o sériovej nekorelovanosti rezíduí je obmedzujúcim faktorom z dôvodu nevýdatnosti odhadov regresných koeficientov. Z tohto dôvodu bol uskutočnený *Breusch-Godfrey Serial Correlation LM Test*, ktorého nulová hypotéza hovorí o neexistencii sériovej korelovanosti rezíduí. Testovacia štatistika NR^2 (počet pozorovaní krát index determinácie) má za platnosti H_o asymptotické chí-kvadrát rozdelenie. Funkcia rozdelenia hustoty pravdepodobnosti "F-statistic“ nie je známa.

Tab.č.4

F-statistic	0.717914	Probability	0.597147
Obs*R-squared	4.343923	Probability	0.361457

Tabuľka číslo 4 znázorňuje, že nemôžeme zamietnuť nulovú hypotézu, tzn. medzi rezíduami nie je sériová korelovanosť. Test bol uskutočnený až po štyri oneskorené premenné.

Existencia heteroskadasticity, prejavujúcej sa zvyčajne rastom rozptylu rezíduí v čase, má za následok zníženie vypovedacej schopnosti odhadov štandardných chýb pri metóde OLS a teda rovnako aj t-testy významnosti vysvetľujúcich premenných sú nespoľahlivé. Vyskytuje sa zvyčajne pri prierezových dátach. Na jej detekciu sme použili *White's Heteroskedasticity Test (White, 1980).* Nulová hypotéza testu hovorí o existencii homoskedasticity rezíduí oproti alternatívnej hypotéze o ich heteroskedasticite. White-ova testovacia štatistika má asymptotické chí-kvadrát rozdelenie s *(k-1)* stupňami voľnosti (k-počet odhadovaných parametrov). Nutným predpokladom testu je normalita, sériová nekorelovanosť náhodných chýb a dostatočne veľký rozsah dát. Tieto predpoklady boli na základe vyššie uvedených testov splnené.

Tab.č.5

F-statistic	0.549761	Probability	0.736323
Obs*R-squared	3.252325	Probability	0.661148

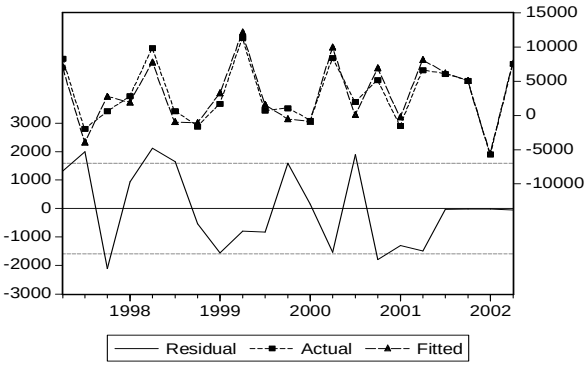
Na všetkých bežných hladinách významnosti nemôžeme zamietnuť H_o.

Existuje však aj jednoduchý, univerzálny a neparametrický (na vplyvné body menej citlivý) test pre zistenie konštantnosti rozptylu (Hušek 1998). Je založený na Spearmanovom koeficiente poradovej korelácie r_s. Kritické hodnoty pre malé *n* sú tabelované (Likeš 1978), pre n > 30 postačuje asymptotický vzťah využívajúci kvantily normálneho rozdelenia*:* $r_s (n-1)^{1/2} = u_{a/2}$. Ani podľa tohto testu nemôžeme zamietnuť na všetkých bežných hladinách významnosti nulovú hypotézu o homoskedasticite rezíduí.

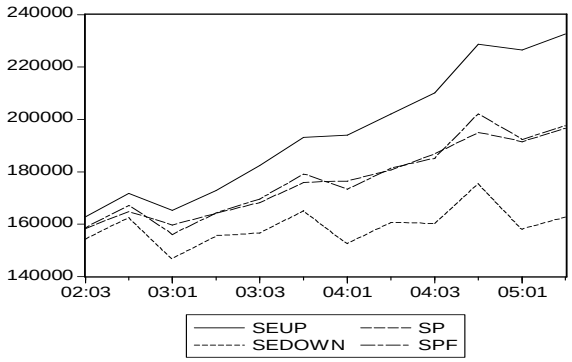
Zamietnutím hypotézy o heteroskedasticite rezíduí bolo ukončené testovanie klasických predpokladov potrebných pre použitie OLS odhadov.

Cieľom tejto prípadovej štúdie bola konštrukcia modelu, ktorý by mal predikčnú schopnosť. Grafickou analýzou, ktorá je znázornená na nasledujúcej strane, sa dá potvrdiť splnenie tohto cieľa. Graf číslo 1 ponúka znázornenie nameraných a vyrovnaných hodnôt spotreby domácností v intervale výpočtu modelu. Znázorňuje taktiež hodnoty rezíduí, ktoré sa podarilo skoro úplne minimalizovať.

Graf číslo 2 znázorňuje “test“ predikčnej schopnosti modelu. Predpoveď bola uskutočnená pre celok dát v období od 2002:3 do 2005:2. Bol zostrojený interval predpovede s hornou a dolnou hranicou, ako “odhadovaná“ hodnota +/- dva krát štandardná odchýlka. S aproximatívne 95 % pravdepodobnosťou sa bude budúca nameraná hodnota závislej premennej nachádzať vo vymedzených hraniciach intervalu. Modelom odhadnuté hodnoty spotreby kopírujú skutočné hodnoty, pričom ani v jednom analyzovanom období sa nepribližujú k vyššie stanoveným hraniciam.



Graf č.1: *Graf skutočných, vyrovnaných a rezíduálnych hodnôt*



Graf č.2: *Predpoveď*

Na základe grafickej analýzy a tiež nízkej hodnoty *Theilovho koeficientu nesúladu* je možné konštatovať, že prezentovaný model má schopnosť poskytovať uspokojivé predpovede.

Zápis modelu

$$SP_t = 1770{,}51 + 0{,}025*(IM_t - IM_{t-4}) + 10{,}87*(MZ_t - MZ_{t-1}) + SP_{t-1} - 0{,}47*u_{t-1} +$$

S.D.	6139,33	0,01	2,33	0,25
t-stat	0,29	3,73	4,66	1,89

$$+ 0{,}92* u_{t-4} - 0{,}43* u_{t-5} + \xi t - 0{,}989* \xi t\text{-}4$$

S.D.	0,09	0,13	
t-stat	10,61	7,65	

kde u_t sú pôvodné rezíduá modelu a $t = 9, \ldots, 42$.

Záver

Cieľ práce, zostrojenie modelu s predikčnou schopnosťou, sa nám podľa vyššie uvedených testov podarilo splniť. Nárast objemu dovezených tovarov a služieb za posledné štyri štvrťroky o jeden milión slovenských korún má za následok zvýšenie konečnej spotreby domácností o 25 tisíc slovenských korún za predpokladu, že ostatné premenné sa nezmenia. Identicky, kladná zmena priemernej mesačnej mzdy oproti predchádzajúcemu štvrťroku o 1 tisíc slovenských korún spôsobí nárast konečnej spotreby domácností o 10,87 milióna slovenských korún, za predpokladu ceteris paribus.

Literatúra

HUŠEK, R. 1999. Ekonometrická analýza. Praha: Ekopress, 1999.
HUŠEK, R. 1998. Základy ekonometrické analýzy, VŠE 1998.
ARTL, J. 1999. Moderní metody modelování ekonomických časových řad, Grada 1999
[b.a.].2002. EViews 4 User´s Guide, Quantitative Micro Software, LLC. 2002

František Štulajter, 4. ročník
Ekonomická fakulta UMB
Tajovského 10
975 90 Banská Bystrica
stulajterf@yahoo.com

Modelovanie bankových vkladov v rokoch 1998-2003

Michal Valkovič

Abstrakt
Theme of contribution is econometric model of bank deposits in years 1998- 2003. We use quarter-annual data. Research was focused on relation between money aggregate M2, bank deposits interest rates, average economy wages and GDP, bank deposits. We have found evidence of strong relation between changes of listed variables. We can constructed an econometric model, which explained big part of dependent variable variance.

1. Dáta

Závislou premennou sú bankové vklady podnikov, obyvateľstva a poisťovní v rokoch 1998-2003. Zdrojom dát boli štvrťročné údaje získané z verejne dostupného archívu NBS. Skúmali sme závislosť bankových vkladov od peňažného agregátu (M2), úrokovej miery bankových vkladov (UM), priemerných miezd v národnom hospodárstve (MZDY) a od hrubého domáceho produktu (HDP).Na analýzu dát sme použili štatistický software Eviews.

2. Vzájomné závislosti sledovaných premenných

Prvotným problémom modelu bola multikolinearita nezávislých premenných, preto sme modelovali namiesto hodnôt týchto veličín absolútne zmeny týchto veličín, a teda nezávislou premennou sú **zmeny bankových vkladov**

Tabuľka 1: Korelačná matica vysvetľujúcich premenných

	D_HDP	**D_M2**	**D_MZDY**	**D_UM**
D_HDP	1.000000	-0.309740	0.119102	-0.088411
D_M2	-0.309740	1.000000	0.181956	-0.178767
D_MZDY	0.119102	0.181956	1.000000	0.025336
D_UM	-0.088411	-0.178767	0.025336	1.000000

Hodnoty vzájomnej korelácie medzi jednotlivými nezávislými premennými nadobúdajú prijateľné hodnoty a preto vylučujeme pravdepodobnosť multikolinearity.

3. Ekonometrický model

3.1. Prvotný model

Prvotný model je založený na primárnych predpokladoch vplyvu zmeny úrokovej miery, peňažného agregátu M2, miezd a hrubého domáceho produktu na zmenu bankových vkladov.

Tabuľka 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.071369	0.720526	0.099051	0.9222
D_UM	0.550509	0.672734	0.818315	**0.4239**
D_M2	0.861390	0.052860	16.29582	0.0000
D_MZDY	0.000823	0.000279	2.952881	0.0085
D_HDP	**-0.007162**	0.040813	-0.175490	**0.8627**
R-squared	0.955883	Mean dependent var		10.83623
Adjusted R-squared	0.946079	S.D. dependent var		7.030359
S.E. of regression	1.632511	Akaike info criterion		4.007776
Sum squared resid	47.97167	Schwarz criterion		4.254623
Log likelihood	-41.08943	F-statistic		97.50122
Durbin-Watson stat	1.511937	Prob(F-statistic)		0.000000

Problémy prvotného modelu sú:
a.) príliš vysoké hodnoty probability D_UM a D_HDP, z čoho vyplýva, že nemôžeme zamietnuť nulovú hypotézu o nulovej hodnote koeficientov týchto nezávislých premenných b.) ako aj záporná hodnota koeficientu HDP.

3.2. Variantný model

Skúmali sme možnosti využitia logaritmickej transformácie a oneskorených nezávislých premenných D_HDP a D_UM a úrokovej miery, aby sme mohol zamietnuť hypotézu H_0 o ich nulových koeficientoch. Zároveň sme použili umelú premennú na vysvetlenie odchýlky zmeny bankových vkladov v prvom štvrťroku roku 1999, na čo mohli mať aj vplyv zmeny politickej moci v krajine. Na základe testov koeficientov sme prijali hypotézu o nulovom koeficiente nezávislej premennej D_HDP.

Tabuľka 3 Odhady variantného modelu (2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.827204	0.413653	1.999754	0.0618
D_UM(-1)	1.536926	0.393739	3.903411	0.0011
D_M2	0.803952	0.027679	29.04535	0.0000
D_MZDY	0.001125	0.000156	7.226501	0.0000
UMPREM	3.529737	1.093979	3.226513	0.0050
R-squared	0.986613	Mean dependent var		11.24091
Adjusted R-squared	0.983463	S.D. dependent var		6.916187
S.E. of regression	0.889405	Akaike info criterion		2.800189
Sum squared resid	13.44771	Schwarz criterion		3.048153
Log likelihood	-25.80207	F-statistic		313.2133
Durbin-Watson stat	1.953881	Prob(F-statistic)		0.000000

Estimation Equation:
==========

$$D\_BV = C(1) + C(2)*D\_UM(-1) + C(3)*D\_M2 + C(4)*D\_MZDY + C(5)*UMPREM + \varepsilon_t \quad (2)$$

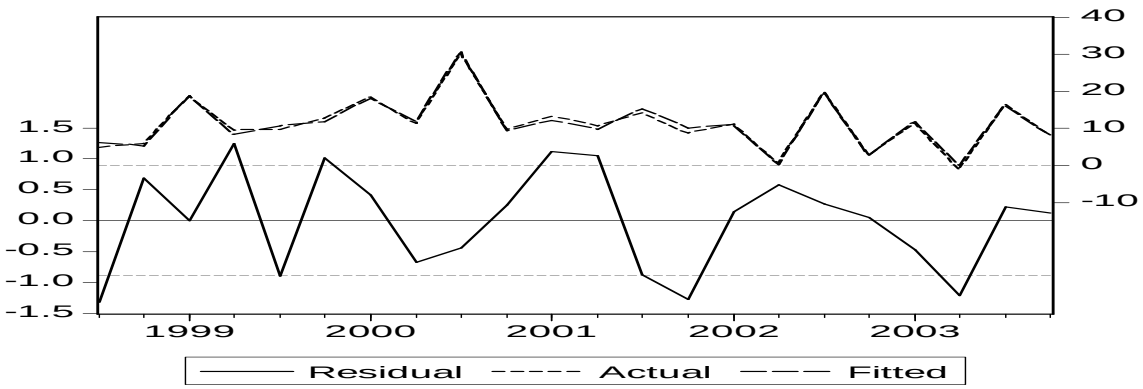
Substituted Coefficients:
==========

$$est(D\_BV) = 0.827 + 1.537*D\_UM(-1) + 0.804*D\_M2 + 0.001*D\_MZDY + 3.53*UMPREM$$

kde,
D_BV - zmena bankových vkladov v mld .SKK,
D__UM(-1) - zmena úrokovej miery bankových vkladov v predchádzajúcom období ($UM_{t-2} - UM_{t-1}$),
D_M2 - zmena peňažného agregátu M2 v mld. SKK,
D_MZDY- zmena miezd v národnom hospodárstve v SKK,
UMPREM- umelá premenná, ktorá má hodnotu 1 v prvom štvrťroku roka 1999, 0 inde.

Umelá premenná v prvom štvrťroku 1999 nám poukazuje na nadmerné zvýšenie bankových vkladov o 3.53 miliardy nad predpokladanú zmenu, čo bolo spôsobené pravdepodobne zmenou vlády v krajine.
Vo všetkých sledovaných štatistických ukazovateľoch došlo k zlepšeniu, nenašli sme problém ani pri ekonomickej verifikácií.

4. Testovanie modelu



Graf 1 Skutočné dáta modelu a rezíduá variantného modelu

Z grafu 1 je zrejmé, že model veľmi dobre kopíruje skutočné hodnoty zmeny bankových vkladov.

4.1. Test normality rezíduí:

Podľa testu normality (p-hodnota = 0,57) nemôžeme zamietnuť nulovú hypotézu o normalite rozdelenia rezíduí. Podľa hodnoty momentového koeficientu špicatosti γ_2=1,953364 je zrejmé, že nami sledované reziduá majú plochšie rozdelenie ako je normované normálne rozdelenie.

4.2. Breusch-Godfreyov LM test sériovej korelácie

Tabuľka 4

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.677493	Probability	0.522767
Obs*R-squared	1.822668	Probability	0.401988

Nulová hypotéza hovorí o neexistencii sériovej korelácie, na základe hodnoty p= 0,5228 nemôžeme na hladine významnosti 0,05 zamietnuť nulovú hypotézu.

4.3. Ramseyov RESET Test

Test správnosti špecifikácie modelu Ramsey RESET test hľadá problémy v reziduách pri testovaní nulovej hypotézy, že stredná hodnota vektora chýb odhadu je rovná nulovému vektoru. Na základe výsledkov testu nemôžeme na všetkých bežných hladinách významnosti zamietnuť hypotézu H_0: E[$\boldsymbol{\varepsilon}$] = **0**

4.4. Test Heteroskedasticity

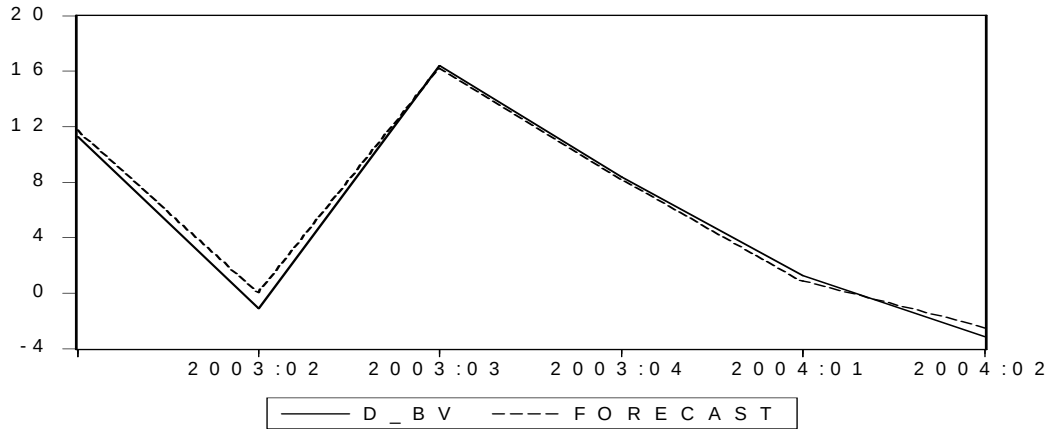
Tabuľka 5

White Heteroskedasticity Test:

F-statistic	1.229812	Probability	0.350053
Obs*R-squared	8.376915	Probability	0.300534

Na základe výsledkov Whiteovho testu nemôžeme zamietnuť nulovú hypotézu o homoskedasticite náhodných chýb.

5. Predikcia



Graf 2 Predikcia

Theilov koeficient nesúladu dosahuje hodnotu 0,034. Predikcia bola vykonávaná na šiestich štvrťrokoch (2003:1-2004:2) a ako je zrejmé z grafu a aj podľa Theilovho koeficientu, predikované dáta nedosahujú významnejšie odchýlky od skutočných dáta a preto môžeme konštatovať, že ekonometrický model je vhodný aj na predikciu

Záver

$$D\_BV = 0.827 + 1.537*D\_UM(-1) + 0.804*D\_M2 + 0.001*D\_MZDY + 3.53*UMPREM + \varepsilon_t \quad (2)$$

Výsledný model potvrdil apriórne hypotézy o priamej závislosti zmien úrokovej miery, miezd a peňažného agregátu M2 na veľkosť zmeny bankových vkladov. Uvedené závislosti sú zhodné s predpokladom osvojeným si pri zostavovaní regresného modelu, ktorý vychádzal zo všeobecnej ekonomickej teórie.
Z ekonometrického modelu vyplýva, že v sledovanom období za predpokladu nezmenených hodnôt všetkých nezávislých premenných, môžeme očakávať v priemere štvrťročný nárast bankových vkladov o 0,827 miliárd SKK. Pri nezmenených ostatných nezávislých premenných v sledovanom období vyvolá rast úrokovej miery o jeden percentuálny bod rast bankových vkladov o 1,537 miliardy SKK, rast peňažného agregátu M2 o jednu miliardu SKK by vyvolal v priemere rast bankových vkladov o 0,804 miliardy SKK za predpokladu ostatných premenných nezmenených. Za predpokladu platnosti modelu v budúcnosti, by rast priemerných miezd v národnom hospodárstve o 1 000 SKK vyvolal v priemere zvýšenie bankových vkladov o 1 mld. SKK.

Zoznam použitej literatúry

HUŠEK, R., PELIKÁN, J.2003. Aplikovaná ekonometrie. Teorie a praxe. Praha : Proffesional Publishing, 2003.
HUŠEK, R. 1999. Ekonometrická analýza. Praha: Ekopress, 1999.

Michal Valkovič, 4. ročník
Ekonomická fakulta UMB
Tajovského 10
975 90 Banská Bystrica
misisiak@orangemail.sk, michal.valkovic@gmail.com

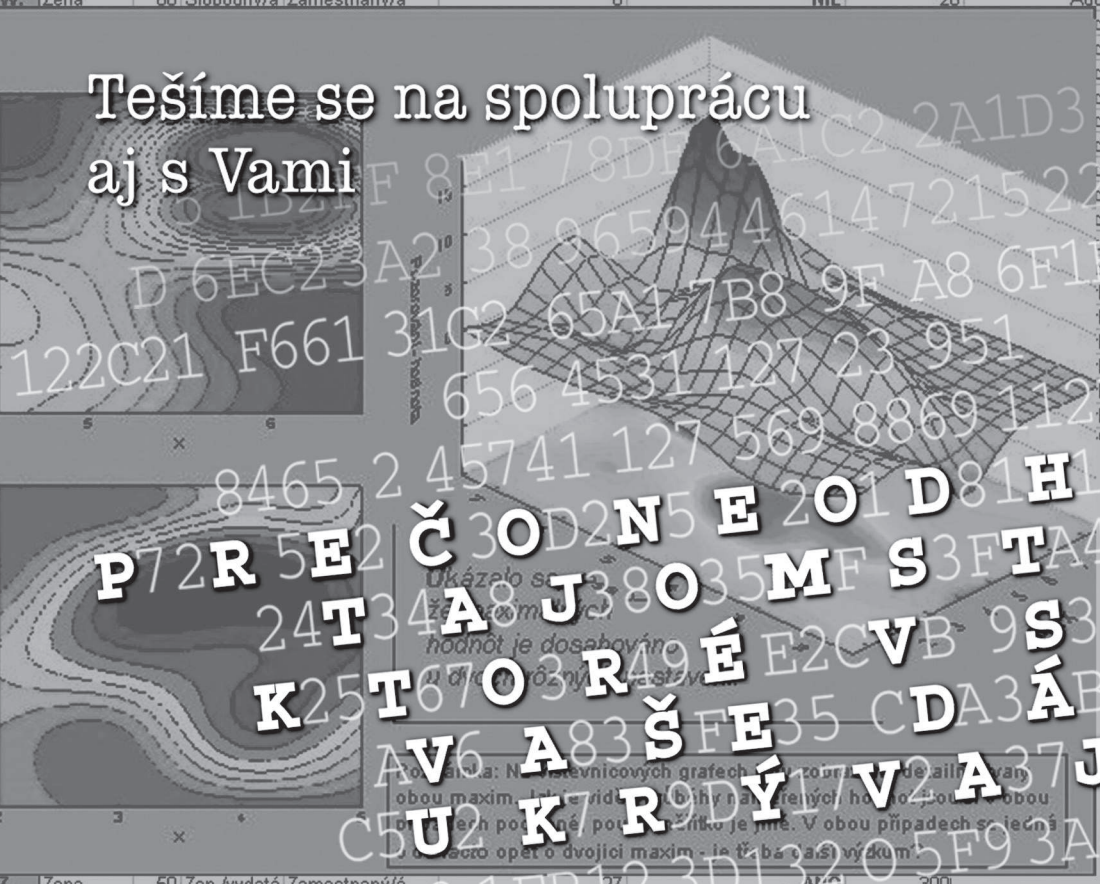
OBSAH

STATISTICA®

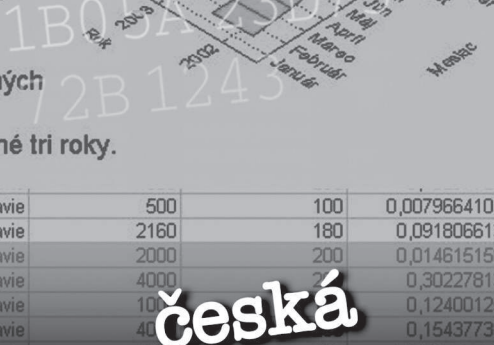
analýza dát • dáta mining • riadenie kvality • analýza cez web

Softvérové riešenia *STATISTICA* sa využívajú v bankách • v poisťovniach • v priemysle vo výskume • na školách • v službách v štátnej správe a inde

Tešíme se na spoluprácu
aj s Vami



PREČO NEODHALIŤ
KTO RÉ V STVÁ,
VAŠE DÁTA
UKRÝVAJÚ?



StatSoft®

česká verzia

StatSoft CR, Podbabská 16, CZ-160 00 Praha 6
www.StatSoft.sk, info@StatSoft.cz, tel. +420 233 325 006