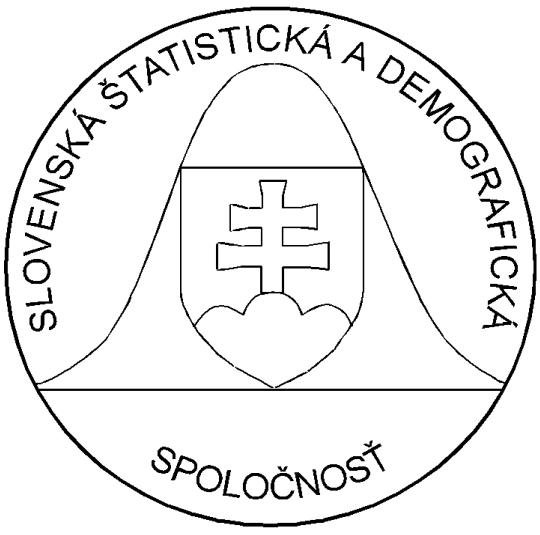
1/2006

FORUM STATISTICUM SLOVACUM



ISSN 1336-7420

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

Naše najbližšie akcie:

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

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

VÝPOČTOVÁ ŠTATISTIKA 2006,
15. Medzinárodný seminár
tematické zameranie: *Praktické využitie paketov štatistických programov, práca s rozsiahlymi súbormi údajov, vyučovanie výpočtovej štatistiky a príbuzných predmetov, praktické aplikácie výpočtovej štatistiky, iné*
7. - 8. 12. 2006, Bratislava

PREHLIADKA PRÁC MLADÝCH ŠTATISTIKOV A DEMOGRAFOV 2006,
uskutočňuje sa v rámci seminára Výpočtová štatistika
7. 12. 2006, Bratislava

KONFERENCIA POHĽADY NA EKONOMIKU SLOVENSKA 2007,
tematické zameranie: *Makroekonomický vývoj; zamestnanosť*
27. 3. 2007, Bratislava

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,
prvé číslo druhého ročníka časopisu, ktorý vydáva Slovenská štatistická a demografická spoločnosť (SŠDS) je zostavené z príspevkov, ktoré autori pripravili pre 13. Slovenskú štatistickú konferenciu. Táto akcia sa uskutočila v dňoch 3. až 5. mája v Malackách. Publikácia obsahuje aj materiály na Valné zhomaždenie členov SŠDS, ktoré sa uskutočnilo v rámci konferencie dňa 3. mája 2006.

Akciu, z poverenia Výboru SŠDS, zorganizoval Organizačný a programový výbor: Ing. Magdaléna Šipková – predsedníčka, RNDr. Ján Luha, CSc. – tajomník, Ing. Silvia Szabová, prof. Ing. Jozef Chajdiak, CSc., RNDr. Peter Mach, 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.

Výbor SŠDS

HLAVNÉ ÚLOHY
Slovenskej štatistickej a demografickej spoločnosti
na obdobie rokov 2006 - 2010

Poslanie a hlavné úlohy Slovenskej štatistickej a demografickej spoločnosti sú sformulované v článku 3 Stanov:

3

Poslanie a hlavné úlohy Spoločnosti

Poslaním Spoločnosti je najmä:

1. Rozvíjať štatistickú a demografickú vedu a jej spoločenské využitie na prospech Slovenskej republiky.
2. Rozširovať poznatky z oblasti štatistiky a demografie.
3. Organizovať konferencie, semináre, sympózia, prednášky, diskusie, odborné školenia a iné akcie.
4. Poskytovať pomoc svojim členom pri ich vedeckej a inej odbornej práci.
5. Zvyšovať odbornú úroveň svojich členov s osobitným zreteľom na mladých pracovníkov.
6. Prezentovať výsledky štatistickej a demografickej vedy v Slovenskej republike na domácich a medzinárodných fórach.
7. Rozvíjať vlastnú publikačnú a edičnú činnosť.

K naplneniu tohto poslania v období 2006 až 2010 budeme realizovať nasledujúce Hlavné úlohy:

KONFERENCIE:

1. Budeme pokračovať v organizovaní Slovenskej štatistickej konferencie a Slovenskej demografickej konferencie ako najdôležitejších akcií SŠDS:

13. Slovenská štatistická konferencia, „Štatistické metódy“, rok 2006, Malacky
14. Slovenská štatistická konferencia, „Regionálna štatistika“, rok 2008, Žilinský kraj
15. Slovenská štatistická konferencia, „5 rokov SR v EU“, rok 2010, Smolenice
16. Slovenská štatistická konferencia, rok 2012, Bansko-Bystrický kraj

11. Slovenská demografická konferencia, „Migrácia“, rok 2007, Košický kraj
12. Slovenská demografická konferencia, „Využitie GIS v demografii“, rok 2009, Trenčianský kraj
13. Slovenská demografická konferencia, „Mesto a vidiek“, rok 2011, Nitriansky kraj
14. Slovenská demografická konferencia, rok 2013, Prešovský kraj

2. V marci roku 2008 zorganizovať „Slávnostnú konferenciu - 40. rokov SŠDS)

3. Budeme každoročne (apríl) pokračovať v organizovaní jednodennej konferencie v Bratislave “Pohľady na ekonomiku Slovenska“.

4. V spolupráci s Českou štatistickou spoločnosťou budeme pokračovať v organizovaní spoločnej česko-slovenskej konferencie STAKAN (2007) resp. slovensko-českej konferencie PRASTAN (2009).

5. Budeme každoročne pokračovať v organizovaní seminára Výpočtová štatistika, týždňovej školy štatistiky Ekomstat, konferencie FERNSTAT.

6. Budeme každoročne pokračovať v organizovaní Prehliadky prác mladých štatistikov a demografov.

7. Budeme sa priebežne podieľať na organizovaní „Demografického popoludnia“.

8. V rámci pobočiek budeme aj naďalej organizovať príslušné „Regionálne popoludnia“ a ďalšie akcie.

9. Priebežne budeme organizovať prednášky a ďalšie akcie.

PUBLIKAČNÁ ČINNOSŤ

Príspevky z jednotlivých konferencií budeme publikovať v Zborníkoch akcií (Pohľady na ekonomiku Slovenska, iné) alebo v našom vedecko-odbornom periodiku FORUM STATISTICUM SLOVACUM

MEDZINÁRODNÁ SPOLUPRÁCA

V svojej činnosti budeme spolupracovať so štatistickými a demografickými spoločnosťami hlavne z krajín z okolia Slovenskej republiky. Spolupráca s Českou štatistickou spoločnosťou a Českou demografickou spoločnosťou má tradičný charakter. Budeme rozširovať spolupracovať v rámci stredoeurópskeho regiónu.

VNÚTORNÁ SPOLUPRÁCA

V rámci Slovenskej republiky budeme spolupracovať s Radou pre vedecké spoločnosti SAV, vedeckými spoločnosťami SAV, vysokými školami, výskumnými pracoviskami, Štatistickým úradom SR, NBS, Infostatom, firmami z hospodárskej praxe, s vedecko-odbornou aj laickou verejnosťou, štátnou správou a miestnou a regionálnou samosprávou.

Správa o činnosti Slovenskej štatistickej a demografickej spoločnosti za obdobie rokov 2002 - 2006

Činnosť Slovenskej štatistickej a demografickej spoločnosti od ostatného Valného zhromaždenia v mesiaci december 2002 prebiehala v stabilnom rytme.

Výbor celé obdobie pracoval v zložení, ako bol zvolený. Zoznam členov výboru, ich funkcie je zverejnený na našej internetovskej stránke www.ssds.sk

AKCIE

Za najdôležitejšiu činnosť považujeme pokračovanie v organizovaní Slovenskej štatistickej konferencie a Slovenskej demografickej konferencie ako najdôležitejších akcií SŠDS. V roku 2004 sa uskutočnila 12. Slovenská štatistická konferencia v Bardejovských kúpeľoch s tematickým zameraním „Štatistika a integrácia“, a teraz v Malackách prebieha 13. Slovenská štatistická konferencia s tematickým zameraním „Štatistické metódy“.

V roku 2003 sa v Tajove uskutočnila 9. Slovenská demografická konferencia, a v roku 2005 v Smoleniciach 10. Slovenská demografická konferencia s tematickým zameraním „Naša demografia - súčasnosť a perspektívy“.

Každá z uvedených konferencií bola s medzinárodnou účasťou a z každej sme vydali Zborník príspevkov.

Od roku 2000 každoročne organizujeme jednodňovú konferenciu v Bratislave “Pohľady na ekonomiku Slovenska“. Slovenská prognostická špička prezentuje svoje odhady budúceho makroekonomického vývoja SR. Z konferencie vydávame Zborník.

V spolupráci s Českou štatistickou spoločnosťou sme organizovali Spoločnú česko-slovenskú konferencie STAKAN (2003) resp. Spoločnú slovensko-českú konferenciu PRASTAN (2005). Konferencia PRASTAN (PRAvdepodobnosť, ŠTAtistiaka, Numerika) v ostatných rokoch prebiehala ako slovenská akcia.

Pravidelne každoročne organizujeme týždňovú školu štatistiky EKOMSTAT (EKOnoMická ŠTATistika). V roku 2006 prebehne už jej 20. ročník. Každoročne začiatkom decembra prebieha medzinárodný seminár Výpočtová štatistika. V roku 2006 prebehne už jeho 15. ročník. V jeho rámci taktiež organizujeme Prehliadku prác mladých štatistikov a demografov. V rokoch 2004 a 2005 prebehli v Tajove prvé dva ročníky konferencie FERNSTAT (Financie, Ekonomika, Riadenie, Názory). V roku 2006 prebehne už tretí ročník.

Zo všetkých konferencií vydávame Zborník príspevkov akcie. V rokoch 2002 až 2004 sme vydali bulletin Forum Metricum Slovacum. Od roku 2005 vydávame vedeckoodborný časopis Forum Statisticum Slovacum. V roku 2005 sme vydali tri čísla. Materiály nášho Valného zhromaždenia sú v tohtoročnom prvom čísle.

Priebežne sa spolupodieľame na organizovaní „Demografického popoludnia“. Organizujeme prednášky významných štatistikov alebo demografov.

V rámci pobočiek budeme aj naďalej príslušné „Regionálne popoludnia“ a ďalšie akcie.

MEDZINÁRODNÁ SPOLUPRÁCA

V rámci svojej činnosti sme spolupracovali so štatistickými a demografickými spoločnosťami hlavne z krajín z okolia Slovenskej republiky. Spolupráca s Českou štatistickou spoločnosťou a Českou demografickou spoločnosťou má tradičný charakter. Spolupracujeme aj v rámci Stredoeurópskeho regiónu. Zorganizovali sme prednášky predstaviteľov EUROSTAT-u.

SPOLUPRÁCA V RÁMCI SR

V rámci Slovenskej republiky primerane spolupracujeme s Radou pre vedecké spoločnosti SAV, Štatistickým úradom SR, NBS, Infostatom, vedeckými spoločnosťami SAV, vysokými školami, výskumnými pracoviskami, firmami z hospodárskej praxe, s vedecko-odbornou aj laickou verejnosťou, štátnou a miestnou samosprávou. Dôležitá je podpora organizácií a firiem pri materiálnom zabezpečení našich konferencií a seminárov.

VNÚTORNÁ PRÁCA

Spoločnosť má okolo 300 riadnych členov (prihlasujú sa nový členovia, ale nám aj zomierajú členovia). V rámci Spoločnosti máme formalizované sekcie Aplikovanej štatistiky, Matematickej štatistiky, Štatistického riadenia kvality a Demografie. V rámci Spoločnosti fungujú regionálne pobočky v Nitre, Banskej Bystrici, Košiciach a Prešove.

Činnosť našej Spoločnosti je založená na odbornom, nepolitickom základe. Napriek tomu spolupracujeme aj s politickými stranami z koalície aj opozície pri riešení odborných otázok. Významní predstavitelia verejného života preberajú záštitu nad našimi akciami.

SLOVENSKÁ ŠTATISTICKÁ A DEMOGRAFICKÁ SPOLOČNOSŤ
82467 BRATISLAVA, MILETIČOVA 3

STANOVY SPOLOČNOSTI

BRATISLAVA
1993

V záujme rozvoja štatistickej a demografickej vedy na území Slovenskej republiky pôsobí Slovenská štatistická a demografická spoločnosť, ako priama pokračovateľka Slovenskej demografickej a štatistickej spoločnosti pri SAV založenej 28.3.1968. Pod novým názvom pôsobí od 14.3.1990.

1

Postavenie, sídlo a pôsobnosť Spoločnosti

1. Slovenská štatistická a demografická spoločnosť (ďalej v texte stanov len Spoločnosť) je dobrovoľné, výberové združenie vedeckých a iných odborných pracovníkov v oblasti štatistiky a demografie, prípadne iných príbuzných disciplín.
2. Sídlom Spoločnosti je Bratislava.
3. Spoločnosť pôsobí na území Slovenskej republiky.
4. Spoločnosť vyvíja činnosť ako vedecká spoločnosť pri Slovenskej akadémii vied.
5. Spoločnosť v medzinárodných stykoch pôsobí pod názvom Slovak Statistical and Demographical Society.

2

Znak Spoločnosti

1. Znak Spoločnosti tvorí kruh o priemere 6 cm. Vo výške 1,5 cm od spodného okraja je vodorovná os, nad ktorou je zobrazená Gaussova krivka s maximom 3,8 cm. V poli pod Gaussovou krivkou je zobrazený znak Slovenskej republiky. Po obvode kruhu je nadpis SLOVENSKÁ ŠTATISTICKÁ A DEMOGRAFICKÁ SPOLOČNOSŤ. V anglickej verzii SLOVAK STATISTICAL AND DEMOGRAPHICAL SOCIETY.
2. Je dovolené znak úmerne zmenšovať alebo zväčšovať.
3. Grafické zobrazenie znaku Spoločnosti je uvedené v prílohe stanov Spoločnosti.

3

Poslanie a hlavné úlohy Spoločnosti

Poslaním Spoločnosti je najmä:

1. Rozvíjať štatistickú a demografickú vedu a jej spoločenské využitie na prospech Slovenskej republiky.
2. Rozširovať poznatky z oblasti štatistiky a demografie.
3. Organizovať konferencie, semináre, sympózia, prednášky, diskusie, odborné školenia a iné akcie.
4. Poskytovať pomoc svojim členom pri ich vedeckej a inej odbornej práci.
5. Zvyšovať odbornú úroveň svojich členov s osobitným zreteľom na mladých pracovníkov.
6. Prezentovať výsledky štatistickej a demografickej vedy v Slovenskej republike na domácich a medzinárodných fórach.
7. Rozvíjať vlastnú publikačnú a edičnú činnosť.

4

Členstvo

1. Členovia Spoločnosti sú riadni, čestní a kolektívni.
2. Riadnym členom Spoločnosti sa môže stať vedecký a iný odborný pracovník v odboroch štatistiky alebo demografie alebo príbuzných disciplín.
3. Čestným členom Spoločnosti sa môže stať významný domáci alebo zahraničný vedecký alebo iný odborný pracovník, ktorý sa zaslúžil o rozvoj demografickej alebo štatistickej vedy. Čestné členstvo udeľuje Valné zhromaždenie na návrh Výboru.

4. Kolektívnym členom Spoločnosti sa môže stať právnická osoba (spravidla škola, vedecká alebo iná organizácia). Kolektívneho člena zastupuje jeden ňou poverený reprezentant, ktorý spĺňa podmienky pre riadneho člena.
5. Členstvo vzniká po zaplatení zápisného a členského poplatku dňom registrácie prihlášky Výborom Spoločnosti. Výšku zápisného a členského určuje Výbor Spoločnosti.

5

Práva a povinnosti členov

1. Člen Spoločnosti má právo:

a. voliť a byť volený do všetkých orgánov Spoločnosti;
b. podávať návrhy a hlasovať o návrhoch podaných na schôdzkach Spoločnosti;
c. byť informovaný a zúčastňovať sa na všetkých jej podujatiach;
d. prednostne získavať časopisy a iné publikácie, ktoré Spoločnosť vydáva;

2. Člen Spoločnosti má povinnosť:

a. zachovávať ustanovenia stanov a plniť platné rozhodnutia jej orgánov;
b. zúčastňovať sa na plnení úloh Spoločnosti a obhajovať jej záujmy;
c. platiť členské príspevky.

6

Zánik členstva

1. Členstvo v Spoločnosti zaniká úmrtím, vystúpením alebo zrušením.
2. Riadne členstvo môže byť zrušené najmä:

a. ak správanie alebo činnosť člena je v rozpore s povinnosťami, ktoré sú určené stanovami;
b. pre neplatenie členských príspevkov (napriek upomienkam) za obdobie dlhšie ako dva roky.

3. O zrušení členstva rozhoduje Výbor Spoločnosti.

7

Orgány Spoločnosti

1. Vedúcimi orgánmi Spoločnosti sú:

a. Valné zhromaždenie
b. Výbor.

2. Kontrolným orgánom Spoločnosti pre hospodársku činnosť je Revízna komisia.

8

Valné zhromaždenie

1. Najvyšším orgánom Spoločnosti je Valné zhromaždenie. Riadne Valné zhromaždenie zvoláva Výbor najmenej raz za štyri roky. Mimoriadne Valné zhromaždenie zvoláva Výbor z vlastnej iniciatívy, alebo na žiadosť aspoň jednej tretiny riadnych členov, a to najneskôr do jedného mesiaca po predložení žiadosti. Valného zhromaždenia má právo zúčastniť sa každý člen Spoločnosti.
2. Do právomoci Valného zhromaždenia patrí:

a. schvaľovať, upravovať a meniť stanovy Spoločnosti;
b. uznášať sa na hospodárskych opatreniach;
c. schvaľovať správy odstupujúceho Výboru a Revíznej komisie;
d. tajným hlasovaním voliť a odvolávať členov Výboru a Revíznej komisie;
e. rozhodovať o odvolaniach proti rozhodnutiu Výboru;
f. voliť čestných členov Spoločnosti;
g. rušiť čestné členstvo Spoločnosti;
h. uznášať sa o zrušení Spoločnosti.

3. Valné zhromaždenie riadi predseda alebo podpredseda alebo iný poverený člen Spoločnosti.
1. Valné zhromaždenie je schopné právoplatne sa uznášať, ak je prítomná najmenej polovica riadnych členov. Ak sa v určený čas nedostaví potrebný počet členov, môže sa po polhodinovej čakacej lehote upustiť od tejto podmienky.
2. Za prijatie uznesenia musí hlasovať nadpolovičná väčšina prítomných členov. Pri rovnosti hlasov rozhoduje hlas predsedu.

9

Výbor Spoločnosti

1. Výbor je výkonným orgánom Spoločnosti. Riadi činnosť Spoločnosti v období medzi Valnými zhromaždeniami.
2. Výbor plní uznesenia Valného zhromaždenia Spoločnosti.
3. Výbor sa skladá z predsedu, podpredsedov, vedeckého tajomníka, hospodára a ďalších členov Výboru. Z titulu svojej funkcie je členom Výboru predseda Štatistického úradu Slovenskej republiky. Valné zhromaždenie môže určiť ďalších členov Výboru z titulu ich funkcií vo významných štatistických alebo demografických pracoviskách. Z titulu svojich funkcií sú členmi Výboru predsedovia pobočiek a sekcií Spoločnosti. Počet členov Výboru určí Valné zhromaždenie s prihliadnutím na celkový počet členstva Spoločnosti. Výbor sa volí na dobu štyroch rokov. Členmi Výboru môžu byť len riadni členovia Spoločnosti.
4. Predsedu, podpredsedov, vedeckého tajomníka a hospodára volí Výbor Spoločnosti spomedzi svojich členov tajným hlasovaním.
5. Za členov Výboru, ktorí trvale prestanú vykonávať svoje funkcie môže Výbor kooptovať nových členov.
6. Schôdze Výboru Spoločnosti zvoláva predseda alebo vedecký tajomník podľa potreby, najmenej dvakrát do roka; prípadne do jedného týždňa po požiadaní aspoň tretiny členov Výboru.
7. Výbor je schopný uznášať sa za prítomnosti aspoň tretiny svojich členov. Za uznesenie Výboru sa považuje návrh, za ktorý hlasovala väčšina prítomných. Pri rovnosti hlasov rozhoduje hlas predsedajúceho.

10

Revízna komisia

1. Dozor nad hospodárskou činnosťou Spoločnosti vykonáva Revízna komisia.
2. Revíznu komisiu volí Valné zhromaždenie z členov Spoločnosti na funkčné obdobie zhodné s funkčným obdobím Výboru. Členovia Revíznej komisie nesmú byť členmi Výboru Spoločnosti. Revízna komisia volí spomedzi svojich členov svojho predsedu. Revízna komisia môže kooptovať nových členov z radov členov Spoločnosti.
3. Členovia Revíznej komisie sa môžu zúčastňovať s hlasom poradným na zasadnutiach Výboru Spoločnosti.
4. Revízna komisia podáva správy o svojej činnosti Valnému zhromaždeniu. Bez jej súhlasu nemôže byť prerokovaná účtovná uzávierka, ani prijaté uznesenie o použití prebytkov alebo úhrade straty.

11

Sekretariát Spoločnosti

1. Sekretariát Spoločnosti tvorí predseda, vedecký tajomník, hospodár, administratívny pracovník a prípadne ďalší členovia Spoločnosti, ktorých určí Výbor.
2. Úlohou Sekretariátu je výkon operatívnych činností Spoločnosti.
3. Práca Sekretariátu podlieha kontrole Výboru.

12

Pobočky

1. Výbor Spoločnosti zriaďuje teritoriálne pobočky.
2. Úlohou pobočiek je plniť úlohy Spoločnosti v okruhu svojej pôsobnosti.

13

Sekcie a odborné skupiny

1. Na lepšie zabezpečenie činnosti zriaďuje Spoločnosť sekcie alebo odborné skupiny, do ktorých sa členovia združujú podľa svojich záujmov a špecializácie.
2. Sekcie a odborné skupiny sa vytvárajú bez ohľadu na teritoriálne členenie Spoločnosti. Podľa potreby sa môžu organizovať aj v rámci pobočiek.
3. Podmienkou členstva v sekcii alebo odbornej skupine je členstvo v Spoločnosti. Člen môže pracovať v ľubovoľnom počte sekcií alebo odborných skupín.
4. Sekciu vedie podpredseda Spoločnosti.

14

Hmotné prostriedky a hospodárenie Spoločnosti

1. Hmotné prostriedky na činnosť Spoločnosti tvoria zápisné, členské príspevky, dotácie poskytované Slovenskou akadémiou vied alebo inými ustanovizňami a príjmy z vlastnej činnosti. Spoločnosť predkladá orgánom SAV výročnú správu o činnosti a hospodárení.

2. Výbor zostaví koncom každého roka plán činnosti a na jeho podklade finančný rozpočet na budúci rok. Okrem toho začiatkom roka zostaví účtovnú uzávierku za uplynulý rok.
3. Spoločnosť hospodári na základe schváleného plánu a rozpočtu na príslušný rok. Platby a príjmy Spoločnosti prebiehajú na účet Spoločnosti. V rámci svojej hospodárskej činnosti Spoločnosť organizuje vedecké, odborné a propagačné akcie s vložným resp. školným a ďalšími poplatkami, realizuje vydávanie a predaj publikácií, v svojich publikáciách a na svojich akciách realizuje propagačnú a reklamnú činnosť za úhradu, organizuje výskumnú, vedeckú, vzdelávaciu a odbornú činnosť za úhradu, realizuje poradenstvo na úseku štatistiky a demografie a príbuzných oblastí za úhradu, zabezpečuje iné činnosti zabezpečujúce príjem prostriedkov na činnosť a chod spoločnosti.
4. Za majetok Spoločnosti zodpovedá Výbor, ktorý správou majetku poverí jedného člena (hospodára).

15

Zastupovanie Spoločnosti

1. V mene Spoločnosti rokujú a podpisujú nasledovní štatutárni zástupcovia: predseda, vedecký tajomník, hospodár. K svojmu podpisu pripojujú názov Spoločnosti a označenie funkcie. Vo veciach hospodársko-finančných sa vyžadujú podpisy dvoch štatutárnych zástupcov Spoločnosti.
2. Výbor môže splnomocniť svojich jednotlivých členov, prípadne aj iné osoby, aby v rozsahu, ktorý určí Výbor zastupovali Spoločnosť a rokovali v jej mene. Osoby takto poverené podpisujú za Spoločnosť s dodatkom “v zastúpení”.

16

Spolupráca

1. Slovenská štatistická a demografická spoločnosť spolupracuje so spoločnosťami, ktoré majú podobné úlohy v Slovenskej republike a v zahraničí.
2. Spoločnosť spolupracuje pri plnení svojich úloh, predovšetkým so Štatistickým úradom Slovenskej republiky, s pracoviskami SAV, vysokými školami, inými vedeckovýskumnými pracoviskami a ďalšími právnickými osobami.

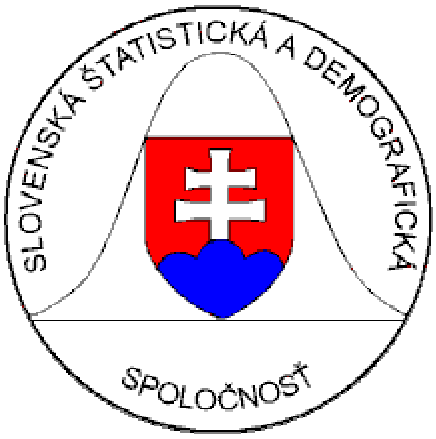
17

Zánik Spoločnosti

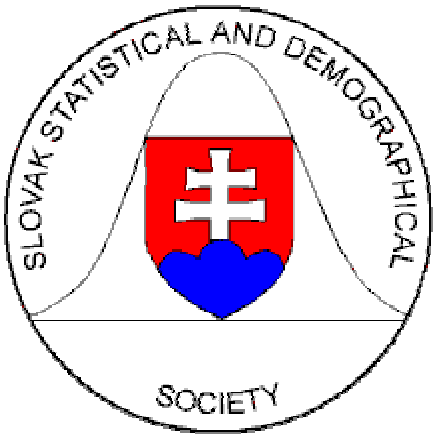
1. Spoločnosť zanikne, ak sa o jej zániku uznesie Valné zhromaždenie dvojtretinovou väčšinou hlasov prítomných členov. Valné zhromaždenie pre účely likvidácie práv a záväzkov Spoločnosti ustanoví likvidačný orgán.
2. O naložení s majetkom, ktorý zostane po úhrade jej záväzkov, rozhodne Valné zhromaždenie.

Príloha
Znak Slovenskej štatistickej a demografickej spoločnosti

a) slovenská verzia



b) anglická verzia



Z histórie Slovenských štatistických konferencií

Prof. Ing. Jozef Chajdiak, CSc. [/1]***, RNDr. Ján Luha, CSc.*** [/2]

[/1] Ekonomická univerzita Bratislava
[/2] Ústav pre výskum verejnej mienky pri ŠÚ SR Bratislava

Slovenská štatistická a demografická spoločnosť mala svoje ustanovujúce zhromaždenie 28. marca 1968. V čase svojho založenia niesla názov Slovenská demografická a štatistická spoločnosť pri SAV. Jej založenie schválilo Predsedníctvo SAV na svojom 33. zasadnutí dňa 18. 12. 1967. Na ustanovujúcom Valnom zhromaždení 28. 3. 1968 bol zvolený výbor v zložení: predseda Doc. Ing. Milan Kovačka, CSc., prvý podpredseda Ing. Daniel Vojtko, CSc., druhý podpredseda Doc. Ing. Ladislav Ivanka, CSc., vedecký tajomník Ing. Rudolf Krč, CSc. a hospodár Dr. Doval. Ďalšie podrobnosti možno získať v príspevku Slovenská demografická spoločnosť založená od D. Vojtka v Ekonomickom časopise 7/1968 XVI str. 715-716. Činnosť Spoločnosti sa orientovala na organizáciu vedecko-odborných podujatí v oblasti štatistiky a demografie prostredníctvom vedeckých konferencií, odborných seminárov a diskusných popoludní. Ku zmene názvu prišlo na Valnom zhromaždení členov spoločnosti dňa 14. marca 1990, kedy sa tiež členská základňa rozšírila o významnú časť štatistickej obce – o matematických štatistikov. Ku tradičným aktivitám Spoločnosti v oblasti aplikovanej štatistiky, výpočtovej štatistky a demografie sa pridružili činnosti z oblasti matematickej štatistiky. Ďalšou významnou oblasťou v činnosti Spoločnosti je štatistické riadenie kvality.

Spoločnosť sa prezentuje pomerne širokými odbornými aktivitami, Slovenská štatistická konferencia patrí medzi najdôležitejšie.

V príspevku uvedieme stručnú chronológiu štatistických konferencií, ktorých názov sa od roku 1996, kedy bola usporiadaná v poradí už šiesta konferencia, ustálil na: Slovenská štatistická konferencia.

Krátko po vzniku Spoločnosti sa organizovala prvá štatistická konferencia v dňoch 16. až 17. 1. 1969. Konferencia sa uskutočnila v Bratislave na Patrónke. Okrem zakladajúcich členov Spoločnosti sa jej zúčastnil tiež predseda Štátneho štatistického úradu Ing. Kazimour a významný slovenský vedec a pedagóg Prof. Briška. Program konferencie bol zameraný na tri okruhy:

1. **okruh - úlohy štátnej štatistiky na Slovensku** (Ing. Ján Marček, CSc.)
2. **okruh - problémy štatistickej vedy** (Doc. Ing. Róbert Štukovský, CSc.)
3. **okruh - systém výučby štatistiky a štatistikov** (Doc. Ing. Ladislav Ivanka, Cíc).

Ďalšie podrobnosti o konferencii možno nájsť v príspevku I. Celoslovenská konferencia štatistikov od R. Krča v Ekonomickom časopise 3/1969 XVII str. 301 - 304. Podľa záverov prvej konferencie sa druhá konferencia mala konať v druhom polroku 1970.

Druhá štatistická konferencia sa uskutočnila v roku 1978 na VŠE v Bratislave na tému **„Aplikované štatistiky“**, za signifikantnej účasti štatistikov z VŠE Praha. Doc. Kovačka prvýkrát načrtol otázky štatistiky životného prostredia.

3. celoslovenská konferencia štatistikov sa uskutočnila 20. novembra 1985 v Bratislave v Dome ROH na tému **„Sociálna štatistika - nástroj riadenia socialistickej spoločnosti“**. Predsedom organizačného výboru bol Doc. Ing. Jozef Chajdiak, CSc.. Materiály konferencie boli opublikované v Informáciách SDŠS pri SAV, ročník XV - 1985, číslo 2.

4. celoslovenská konferencia štatistikov na tému **„Meranie a analýza ekonomickej efektívnosti“** sa uskutočnila 10. - 12. októbra 1988 v Smoleniciach. Od tohto momentu sa štatistická konferencia usporiadava každé dva (párne) roky a materiály konferencie sa vydávajú v samostatnom Zborníku príspevkov. Predsedom organizačného výboru bol Doc.

Ing. Jozef Chajdiak, CSc.. Konferencia sa uskutočnila s medzinárodnou účasťou. Materiály konferencie boli opublikované v Zborníku príspevkov 4. celoslovenskej konferencie štatistikov.

5. celoslovenská konferencia štatistikov na tému **„Súčasné úlohy štatistiky“** sa uskutočnila 2. - 4. mája 1990 v Smoleniciach. Predsedom organizačného výboru bol Ing. Jozef Brezák. Od tohto roku sa vo významne väčšej miere zúčastňujú na konferenciách matematický štatistici. Materiály konferencie boli opublikované v Zborníku príspevkov 5. celoslovenskej konferencie štatistikov.

6. Slovenská štatistická konferencia na tému **„Slovenská štatistika – súčasnosť a perspektívy“** sa uskutočnila 4. - 6. mája 1992 v Smoleniciach. Predsedom organizačného výboru bol RNDr. Ján Luha, CSc.. Materiály konferencie boli opublikované v Zborníku príspevkov 6. Slovenskej štatistickej konferencie.

7. Slovenská štatistická konferencia na tému **„Štatistika v manažmente“** sa uskutočnila 4. - 5. mája 1994 v priestoroch Ekonomickej univerzity v Bratislave. Predsedom organizačného výboru bol Doc. Ing. Jozef Chajdiak, CSc.. Materiály konferencie boli opublikované v Zborníku príspevkov 7. Slovenskej štatistickej konferencie.

8. Slovenská štatistická konferencia na tému **„Štatistické metódy v bankovníctve, finančníctve a poisťovníctve“** sa uskutočnila 8. - 10. mája 1996 na Donovaloch. Predsedom organizačného výboru bol Ing. Vladimír Úradníček, CSc. Materiály konferencie boli opublikované v Zborníku príspevkov z VIII. slovenskej štatistickej konferencie s medzinárodnou účasťou (ISBN 80 - 967343 - 4 - 2). Táto konferencia bola prvou zo štatistických konferencií, ktoré sa od tohto času organizujú v jednotlivých krajoch Slovenskej republiky.

9. Slovenská štatistická konferencia na tému **„Podniková štatistika“** sa uskutočnila 21. - 23. októbra 1998 v Herľanoch. Predsedom organizačného výboru bol Doc. RNDr. Michal Tkáč, CSc.. Materiály konferencie boli opublikované v Zborníku príspevkov 9. Slovenskej štatistickej konferencie (ISBN 80 - 967658 - 3 - 3).

10. Slovenská štatistická konferencia na tému **„Slovenská štatistika - súčasnosť a perspektívy“** sa uskutočnila 10. - 12. mája 2000 v Smoleniciach. Predsedom organizačného výboru bol Doc. Ing. Jozef Chajdiak, CSc., tajomníkom RNDr. Ján Luha,CSc.. Materiály konferencie sú opublikované v Zborníku príspevkov (ISBN 80 - 88946 - 06 - 9). Konferencia mála charakter jubilea - 10. konferencia v roku 2000.

11. Slovenská štatistická konferencia na tému **„Štatistické metódy v praxi“** sa uskutočnila 11. - 13. septembra 2002 v Nitre pod záštitou predsedu ŠÚ SR RNDr. P. Macha na tému „Štatistické metódy v praxi“. Predsedom organizačného výboru bola Doc. RNDr. Beáta Stehlíková, CSc. (stehliko@uniag.sk). Spoluorganizátorom konferencie bola Fakulta ekonomiky a manažmentu Slovenskej poľnohospodárskej univerzity v Nitre. Materiály konferencie boli publikované v Zborníku príspevkov (ISBN 80 - 88946 - 19 - 0).

12. Slovenská štatistická konferencia sa uskutočnila 4. až 6. októbra 2004 v Bardejovských kúpeľoch, pod záštitou podpredsedu vlády SR Pála Csákyho. Hlavným tematickým zameraním bolo **„Štatistika a integrácia“**. Boli stanovené nasledovné tematické okruhy: integrácia do Európskej únie z pohľadu: metodológie štatistiky, histórie štatistiky, matematickej štatistiky, ekonomickej štatistiky, demografickej štatistiky, štátnej štatistiky, poľnohospodárskej štatistiky, výpočtovej štatistiky, štatistického riadenia kvality, bankovej a finančnej štatistiky, poistnej štatistiky, sociálnej štatistiky, regionálnej štatistiky, medzinárodnej štatistiky, legislatívnej štatistiky, etiky štatistiky, organizačnej štatistiky, zahraničných stykov, výučby štatistiky a iných. Predsedom organizačného a programového výboru bol Ing. Ján Cuper (Jan.Cuper@statistics.sk). Materiály konferencie sú opublikované v Zborníku príspevkov (ISBN 80 - 88946 - 37 – 9).

13. Slovenská štatistická konferencia sa uskutočňuje 3. až 5 mája 2006 v hoteli ATRIUM v Malackách. Predsedom organizačného a programového výboru je Ing. Magdaléna Šipková. Okruhy konferencie sú nasledovné: Aplikácie štatistických metód vo vede a praxi, Teória štatistických metód, Metodológia a prax zberu štatistických údajov, Matematická štatistika a pravdepodobnosť, Štatistický softvér, Iné. Príspevky sú publikované v odbornom časopise SŠDS FORUM STATISTICUM SLOVACUM 1/2006.

Tematické zameranie jednotlivých konferencií bolo ovplyvnené ústrednou témou určenou pre danú konferenciu. Ústredné témy zvolené pre konferencie vyjadrujú snahu organizátorov zachytiť najvýznamnejšie potreby a impulzy prieniku štatistiky a ostatných oblastí vedy a praxe.

Súčasné tematické zameranie „Štatistické metódy“ bolo zvolené v súlade s uvedenými kritériami.

Autori, súc si vedomí problémov s pamäťou, začali zostavovať aspoň chronológiu konferencií. S ohľadom na uvedené budeme vďační všetkým, ktorí nám poskytnú doplňujúce informácie o našich konferenciách na e-mailové adresy:

Jan.Luha@statistics.sk
chajdiak@statis.biz
alebo osobne.

Hlavné charakteristiky Bratislavského kraja

Ing. Magdaléna Šipková
Štatistický úrad - Krajská správa v Bratislave

Bratislavský kraj vznikol na základe zákona NR SR č. 221/1996 Z. z. o novom územnom a správnom usporiadaní Slovenskej republiky. Rozlohou 2 052,6 km^2 je najmenším krajom Slovenskej republiky, keď z jej rozlohy zaberá 4,2 %. Rozprestiera sa v juhozápadnej časti SR, kde vytvára spoločnú hranicu s Rakúskou a Maďarskou republikou. Na severe, východe a juhovýchode hraničí s Trnavským krajom. V kraji je najzápadnejší bod v Slovenskej republike na západnej hranici s Rakúskou republikou v obci Záhorská Ves v okrese Malacky. V rámci stredoeurópskeho regiónu má výhodnú centrálnu polohu, plní funkciu medzinárodnej križovatky v cestnej a železničnej doprave, narastá význam dunajskej vodnej cesty a tiež medzinárodnej leteckej dopravy.

Územie kraja je v západnej časti tvorené Záhorskou nížinou, z juhozápadu na severovýchod sa rozkladá pohorie Malých Karpát, východnú a juhovýchodnú časť tvorí Podunajská nížina. Klimaticky patrí kraj do mierne teplej oblasti, na juhu do teplej. Vodná sieť patrí do povodia Dunaja, druhej najväčšej európskej rieky, ktorá preteká cez územie kraja v dĺžke 37 km a s Rakúskou a Maďarskou republikou tvorí časť prirodzenej hranice kraja.

Z historického hľadiská je územie kraja prepojené s územím Bratislavského komitátu a stredovekými stolicami a župami, najmä Bratislavskou, ktorých hranice sa menili iba málo. Nie je zanedbateľné, že na území kraja ležia 4 bývalé slobodné kráľovské mestá, a to Bratislava (od roku 1291), Modra (od roku 1607), Pezinok (od roku 1647) a Svätý Jur (od roku 1647). Cez ich územie a priľahlé podkarpatské obce viedla známa malokarpatská vínna cesta.

Nerastné bohatstvo tvoria ložiská ropy a zemného plynu v Záhorskej nížine, ktoré majú nadregionálny význam. Významné sú aj suroviny pre výrobu stavebných materiálov ako vápence a dolomity, štrkopiesky a stavebný kameň, sialitické a zlievárenské piesky. Zásoby týchto surovín majú prevažne miestny charakter.

Územie kraja sa člení na 8 okresov: Bratislava I, Bratislava II, Bratislava III, Bratislava IV, Bratislava V, Malacky, Pezinok a Senec. Územie hlavného mesta SR tvoria okresy Bratislava I až Bratislava V. Najväčším okresom v rámci Bratislavského kraja je okres Malacky s rozlohou 949,6 km^2 a s podielom 46,3 % z celkovej rozlohy kraja. Najmenším okresom nielen v rámci kraja ale aj v celej Slovenskej republike s plochou 9,6 km^2 a podielom 0,5 % je okres Bratislava I.

Bratislavský kraj má najmenší počet obcí 73, z ktorých 7 má štatút mesta. Sú to hl. mesto SR Bratislava, Malacky a Stupava v okrese Malacky, Pezinok, Modra a Svätý Jur v okrese Pezinok, Senec v okrese Senec. Podiel všetkých miest na celkovom počte obyvateľov kraja v roku 2005 činil 83,1 %, a podiel hl. m. SR Bratislavy 70,5 %.

Počtom obyvateľov 603 699 k 31.12.2005 sa Bratislavský kraj zaraďuje na 6. miesto v medzikrajskom porovnaní, čo predstavuje 11,2 % z úhrnu SR. Hustota obyvateľstva 294,1 obyvateľov na 1 km^2 je 2,7 krát vyššia ako priemer SR a najvyššia v porovnaní s ostatnými krajmi. Hustota obyvateľstva kraja sa pohybuje od 69,1 obyvateľov na 1 km^2 v okrese Malacky do 4400 obyvateľov na 1 km^2 v okrese

Bratislava I. Vo vývoji obyvateľstva sa dlhodobo prejavuje nepriaznivý vývoj prirodzeného pohybu, keď v roku 2005 oproti predchádzajúcemu roku sa zaznamenal prirodzený úbytok o 0,08 ‰. Migračný prírastok o 4,34 ‰ dosiahnutý vplyvom vyšších prírastkov v okresoch Bratislava II, Bratislava III, Bratislava IV., ale najmä v mimobratislavských okresoch, znamenal celkový prírastok obyvateľov v kraji o 4,26 ‰. Zhoršuje sa vekové zloženie obyvateľstva znižovaním podielu vekovej kategórie 0 -14 ročných (13,2 %) a zvyšovaním podielu osôb v poproduktívnom veku (20,9 %). Tento vývoj dokumentuje, že Bratislavský kraj má degresívnu vekovú skladbu s najvyšším indexom starnutia s hodnotou 158,0 v medzikrajskom porovnaní v roku 2005. Najvyššie indexy starnutia boli v okresoch Bratislava I a Bratislava III, v ktorých počet obyvateľov v poproduktívnom veku bol viac ako 2-násobne vyšší ako počet detí do 14 rokov.

Z pohľadu národnostnej štruktúry bol v kraji výrazne dominantný podiel obyvateľov slovenskej národnosti 91,3 %, kým druhá najpočetnejšia maďarská národnosť predstavovala iba 4,6 % podľa výsledkov SODB 2001. Keďže v kraji je sústredených najviac vysokých škôl a univerzít, je to jeden z dôvodov celkovej vyššej vzdelanostnej úrovne obyvateľov. Na vysokoškolských vzdelaných pripadlo 17 %, čo je o 9,1 bodu viac ako celoslovenský priemer. Z hľadiska náboženského vyznania najviac obyvateľov kraja sa prihlásilo k rímskokatolíckemu vyznaniu 61,9 %, kraj sa vyznačuje aj najvyšším podielom osôb bez vyznania (24,9 %) v rámci krajov.

Organizácie so sídlom v kraji dosahujú z úhrnu Slovenskej republiky v prevažnej väčšine ukazovateľov najvyšší podiel v porovnaní s ostatnými krajmi. Na tvorbe hrubého domáceho produktu a pridanej hodnote sa Bratislavský kraj podieľa približne štvrtinou.

V Bratislavskom kraji ku koncu roku 2005 pracovalo 340 174 zamestnancov (priemerný prepočítaný počet) a oproti minulému roku sa zamestnanosť zvýšila o 1,1 %. Na celkovej zamestnanosti Slovenskej republiky sa kraj podieľal 28,1 %. V kraji je najvyššia úroveň priemernej nominálnej mesačnej mzdy na zamestnanca 25 126 Sk, ktorá je o tretinu vyššia ako priemer za SR. V členení podľa odvetví najvyššiu priemernú mzdu dosiahli zamestnanci finančného sprostredkovania 37 177 Sk a v priemysle spolu 28 797 Sk.

Počet evidovaných uchádzačov o zamestnanie k 31.12.2005 bol v Bratislavskom kraji 9 810 osôb a v porovnaní s rokom 2004 klesol o 2 826 osôb, s podielom 2,9 % z celkovej nezamestnanosti SR. Miera evidovanej nezamestnanosti 2,60 % si v kraji udržuje dlhodobo najnižšiu úroveň, pričom priemer za SR predstavoval 11,36 %.

V štatistickom registri organizácií kraja bolo k 31.12.2005 zaregistrovaných 93 301 ekonomicky aktívnych subjektov , čo je o 3 748 (4,2 %) viac ako v minulom roku. Z uvedeného počtu právnické osoby tvorili 35,8 % a fyzické osoby - podnikatelia nezapísaní v OR 64,2 %. Z 33 411 právnických osôb bolo 25 505 podnikov zameraných na tvorbu zisku a z nich najpočetnejšiu skupinu tvorili obchodné spoločnosti (94,3%). Z hľadiska vlastníckych vzťahov prevažná časť podnikov bola v súkromnom vlastníctve (97,5 %). Z celkového počtu podnikateľských subjektov právnických osôb v SR tvoril Bratislavský kraj 30,5 %. Počet fyzických osôb nezapísaných v OR dosiahol 59 890, z ktorých 94,7 % boli živnostníci. Z úhrnu SR dosiahol 16,3 %.

Použitá literatúra:
1. Bulletin 4/2005 ŠÚ SR - Krajská správa v Bratislave
2. Okresy Bratislavského kraja 2004

Malacky – dobré miesto pre život

Príhovor primátora RNDr. Jozefa Ondrejku

Tak ako je vzácna ktorákoľvek ľudská bytosť, a to už len svojou neopakovateľnosťou, tak analogicky - keď hovoríme o mestách - každé z nich je vzácne svojou jedinečnosťou. Má svoje perly, ktoré ho krášlia a vytvárajú náhrdelník svojráznosti. Čím sú osobité Malacky?

Netreba chodiť do Ríma, stačí prísť do Malaciek

Historických pamiatok máme síce ako šafranu, o to sú však cennejšie. Malačania sú pyšní na Zámocký park, kaštieľ, Synagógu a v neposlednom rade na Františkánsky kláštor a kostol. Jeho súčasťou sú Sväté schody - schody, po ktorých v Jeruzaleme kráčal Ježiš do súdnej siene Pilátovej. V 4. storočí ich previezli do Ríma a v Malackách existuje ich verná kópia. Preto ten, kto chce vidieť Sväté schody, nemusí chodiť do Ríma - stačí ak príde do Malaciek.

Prasa v Malackách – to je omyl

Jedným zo symbolov nášho mesta je erb, v ktorom je dominantne zobrazené divé prasa. Jeho tvorca pravdepodobne mylne vychádzal z predpokladu, že názov mesta je spätý s týmto zvieraťom (malacz- v maďarčine znamená prasa). Podľa zistení historikov je však názov mesta odvodený od pomenovania potoka Malina, ktorý mestom tečie. Sofizmom by mohlo byť tvrdenie, že ak v Malackách uvidíte prasa, tak je to určite iba omylom.

Pozor na vtipy o Záhorákoch

Záhoráčtina je kodifikovaný jazyk a je len škoda, že unikol pozornosti pána Štúra. Mohol byť spisovým jazykom všetkých Slovákov a všetci sme mohli ťažiť z jeho nepochybných kvalít a predností - záhoráčtina napríklad nepozná tvrdé y. Takto teraz náš región a ostatné Slovensko spájajú len vtipy o Záhorákoch. Ozaj, viete prečo ich tak radi rozprávame? Je to vďaka Einsteinovej teórii relativity. Z nášho pohľadu sú totiž *Záhorákmi - ľuďmi žijúcimi za horami, Malými Karpatami* - práve všetci ostatní obyvatelia Slovenska! Preto pozor, stojí naozaj za uváženie, koho sa tie vtipy týkajú. Ale, samozrejme, nielen z humoru sme živí.

Malacky – mesto budúcnosti

Ambíciou Malaciek, hádam ako každého iného mesta, je vytváranie lepších podmienok pre život. Aj preto si Malacky zvolili stratégiu dostatočne zvýrazňovať svoje silné stránky, vytvárať priestor pre ich rozvoj tak, aby mesto bolo čo najviac atraktívne a príťažlivé. Takáto gravitácia koncentruje záujem o mesto, vytvára potenciál a predpoklady na elimináciu tých jeho slabých stránok, na ktoré by inak nikdy nebolo dosť peňazí. Výsledkom sú projekty priemyselno-technologického parku Eurovalley, inkubátora pre malé a stredné podnikanie či pozoruhodné zníženie nezamestnanosti - za ostatné štyri roky takmer trojnásobné....

Osobnosti mesta

Mesto, to nie sú len projekty, budovy, chodníky a cesty. Mesto sú predovšetkým ľudia. Vďaka nim žije, kvôli nim má význam plánovať budúcnosť. K Malackám sa hlásia významní rodáci roztrúsení po celom svete, ale aj súčasné osobnosti mesta. Nie je tu dostatok priestoru, aby som vymenoval všetky žijúce osobnosti Malaciek. Podľa mňa ich je asi 18 tisíc. Spoločne sa podieľajú na vytváraní obrazu nášho mesta. Spolu si z Malaciek tvoríme dobré miesto pre život. A ak vás niektorá perla z náhrdelníka zvaného Malacky upútala, budeme radi, ak naše mesto opäť navštívite.

MESTO MALACKY

Mesto leží v centre južnej časti Záhorskej nížiny, 35 km od Bratislavy smerom na Brno. Západne od mesta tečie rieka Morava, východne sa tiahnu Malé Karpaty.



Prvá písomná zmienka o meste sa nachádza na darovacej listine uhorského kráľa Ondreja II. z roku 1206. V roku 1622 sa na tristo rokov stal majiteľom mesta rod Pálfiovcov. K historickým a kultúrnym pamiatkam mesta patrí pálfiovský rodový kaštieľ (1624), rozsiahly Zámocký park (anglický park so vzácnymi drevinami), františkánsky kláštor s ochranným múrom a štyrmi vežičkami (1653), Chrám Nepoškvrneného Počatia Panny Márie (nachádza sa v ňom aj svetový unikát: kópia jeruzalemských Svätých schodov z roku 1653), farský kostol Najsvätejšej Trojice (1604), židovská synagóga (1886) a stála expozícia Krypty pod Františkánskym kláštorom. Malacky sú dnes mestom s bohatými kultúrnymi a športovými tradíciami. Mesto pripravuje tradičné happeningové podujatia Pravomájové Malacky, Kultúrne leto v Zámockom parku a Adventné Malacky. Možnosti oddychu a rekreácie: letné kúpalisko s 50- metrovým plaveckým a detským bazénom, športová hala Malina s krytou plavárňou, fitnesscentrom, kolkárňou a saunou. V blízkosti mesta vyrastá Priemyselno-technologický park Záhorie (Eurovalley).

Ďalšie informácie:
Športová hala Malina , Sasinkova 2, Malacky, 034/7723 202
Mestské centrum kultúry, Záhorácka 1919, Malacky, 034/7722 155

Adresa:
Radlinského 1
901 01 Malacky

Tel.: 00421/34/79 66 100
Fax: 00421/34/79 66 151
www.malacky.sk
e- mail: msu@malacky.sk



Rekonštruovaný Zámocký park je dejiskom celomestských happeningových podujatí. V septembri sa tu uskutočnia historické Osemstoročné Malacky.



Pálfiovský kaštieľ - perla Zámockého parku v Malackách.



Kostol Nepoškvrneného Počatia Panny Márie v Malackách.



Hlavný oltár kostola je najcennejšou cirkevnou pamiatkou mesta. Vrchol oltára je zakončený ozdobnou kartušou s erbom Mikuláša Pálfiho a jeho manželky Kataríny Weichsovej.

Podmienky aproximovateľnosti binomického rozdelenia normálnym rozdelením a stanovenie rozsahu výberu

Ján Luha

Abstract
Arcticle dealt with conditions for aproximation binomial distribution by means of normal distribution and determination of sample size.

Úvod

Pri výberových štatistických zisťovaniach sa často stretávame s kvalitatívnymi znakmi, pričom je úlohou získať reliabilné odhady relatívnych frekvencií, čo súvisí s rozsahom výberu. Obmedzíme sa na najjednoduchšie znaky s dvomi možnými odpoveďami – vtedy ide o skúmanie podielov. Za predpokladu, že výberové zisťovanie je realizáciou náhodného výberu je rozdelenie sledovaných znakov binomické. Pri štatistických odhadoch a testoch sa často využíva možnosť aproximácie normálnym rozdelením, pritom však niektorí autori nesledujú podmienky aproximovateľnosti a vzťahy pre normálne rozdelenie používajú aj keď tieto podmienky neplatia.

Podmienky aproximovateľnosti

Uvažujme teda kvalitatívny znak X, ktorý vyjadruje prítomnosť skúmaného javu v populácii. Podiel (pravdepodobnosť) výskytu skúmaného javu budeme označovať symbolom π a odpovedajúci výberový odhad ako p.
Nech n je rozsah výberového súboru a x je početnosť sledovaného variantu kvalitatívneho štatistického znaku. Potom **podielom** je veličina: **p=x/n**.
Veličina p má binomické rozdelenie s parametrami n, π.

Podmienky aproximovateľnosti rozdelenia pravdepodobnosti veličiny p pomocou normálneho rozdelenia sa vzťahujú na „veľkosť“ hodnoty podielu π a na „veľkosť“ rozsahu výberu.

a) Podmienky na „veľkosť“ podielu:

Uvazujme funkciu pravdepodobnosti binomického rozdelenia B(n, π):

$$P(X=k)=\binom{n}{k} \pi^k (1-\pi)^{n-k}$$

Táto funkcia je:
1. pre $\pi < 1/ (1/1+n)$ klesajúcou funkciou k,
2. pre $\pi > 1/ (1/1+n)$ rastúcou funkciou k a
3. pre $\pi \in \{1/(n+1), n/(n+1)\}$ je najprv rastúca a potom klesajúca.

Z uvedených vlastností a podľa limitných viet teórie pravdepodobnosti dostávame vzťah 3. ako podmienku aproximovateľnosti vzľadom na „veľkosť“ podielu.
Pozri napr.: Luha J.: Testovanie štatistických hypotéz pri analýze súborov charakterizovaných kvalitatívnymi znakmi. STV Bratislava, 1985, str. 54.

b) Podmienky na „veľkosť“ rozsahu výberu:

Minimálny potrebný rozsah možno stanoviť na základe podmienky

$$n.p.q > 9 \quad .$$

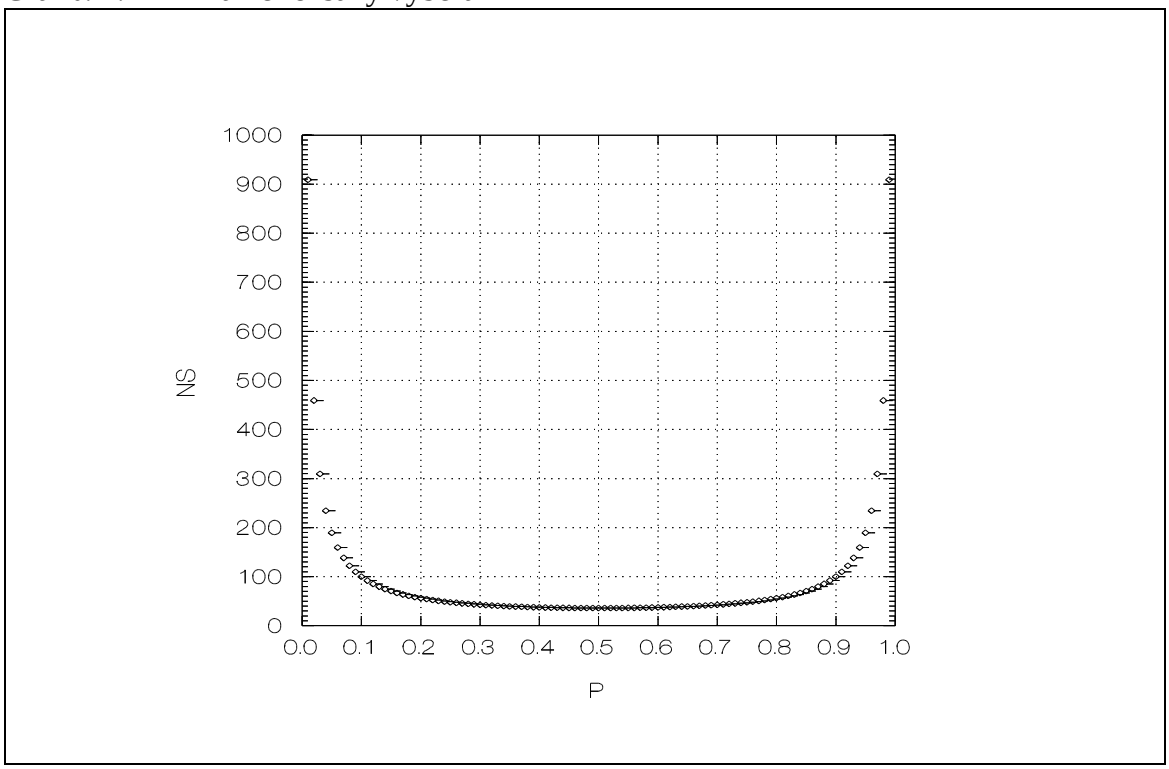
Pozri napr. Hald A.: Statistical Theory with Engeneering Applications. Wiley, New York 1952 paragraf 21.5, str.680.
Približnú orientáciu o potrebnom rozsahu výberu dáva tabuľka č. 1:

Tabuľka č.1: Minimálne rozsahy 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**

Graficky sme množinu minimálnych hodnôt rozsahu výberu zobrazili na grafe č.1. Minimálny rozsah výberu získame z hodnôt nad zobrazenou krivkou, kde na osi sú y hodnoty minimánleho rozsahu výberu , ktorý sme označili NS a na osi x sú hodnoty podielu P .

Graf č.1 : Minimálne rozsahy výberu



Stanovenie rozsahu výberu

Teoretickým východiskom pre stanovenie rozsahu výberovej vzorky, vzhľadom na to, že otázky dotazníka sú kvalitatívne, pri jednoduchom náhodnom výbere je vzorec:

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

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:

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

kde f=n/N je výberový pomer.

Na konkretizáciu výpočtov potrebujeme skúmať očakávanú presnosť (reliabilitu) výsledkov a potom na základe rozboru realizovateľnosti prieskumu zvoliť konkrétnu hodnotu rozsahu výberov.

Obvykle sa vzorec pre výpočet minimálneho rozsahu výberu zjednodušuje na prípad p=0,5. Potom dostávame: $\mathbf{n=(0,98/d)^2}$.

Uvádzame tabuľku minimálnych rozsahov výberu pre vybrané hodnoty požadovanej presnosti (reliability):

reliabilita d	0,01	0,02	0,03	0,05	0,08	0,10	0,11	0,12	0,1265	0,15
rozsah n	9064	2401	1067	384	150	96	79	67	60	43

Pritom však musíme sledovať splnenie podmienok aplikácie uvedeného vzorca.

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. Chajdiak J.: Štatistika jednoducho. Statis Bratislava 2003.
4. Hald S.: Statistical Theory with Engeneering Applications. Wiley, New York 1952.
5. Johnson N.L., Kotz S. Distributions in Statistics. Discrete Distributions. Wiley, New York 1969.
6. Luha J.: Testovanie štatistických hypotéz pri analýze súborov charakterizovaných kvalitatívnymi znakmi. STV Bratislava, 1985.
7. Luha J.: Exaktné intervaly spoľahlivosti pre podiely. Slovenská štatistika a demografia 1/96.
8. Luha J.: Meranie spoľahlivosti výskumov verejnej mienky pomocou intervalov spoľahlivosti pre podiely. FORUM STATISTICUM SLOVACUM 1/2005. SŠDS Bratislava 2005.
9. Luha J.: Reprezentatívnosť vo výskumoch verejnej mienky. FORUM STATISTICUM SLOVACUM 2/2005. SŠDS Bratislava 2005.
10. Motulsky H.: GraphPad PRISM, Version 4.0 Statistical Guide. Statisstical analyses for laboratory and clinical research. 2005. GraphPad Software, Inc. http://www.graphpad.com/

Kontakt na autora:
RNDr. Ján Luha, CSc.
Jan.Luha@statistics.sk

Rozhodovanie s viacerými cieľmi a Taguchiho metóda

Milan Terek

Abstract. The article deals with one possibility of the multiattribute decision making in the Taguchi method. The method of the using of the Design of Experiments in the robust design is sometimes called Taguchi method. When we consider more design attributes, the configuration of factors and their levels preferred according to one attribute, does not must be preferred also according to other attributes. The determination of some common valuation criteria is necessary. One from the possible approach is the utility function method in which the individual utility functions are aggregated to the multiattribute utility function. We consider additive utility function. The possibility of the application is illustrated on the example.

Úvod

Program nepretržitého zlepšovania kvality spočíva podľa G. Taguchiho v redukcii variability *výkonových charakteristík produktov* (výkonové charakteristiky produktu sú ukazovatele kvality produktu, ktoré determinujú jeho výkon z hľadiska miery uspokojenia požiadaviek používateľov produktu) vzhľadom na ich *cieľové hodnoty* (*target values – T*)[1].

Pri hľadaní a analýze alternatívnych konfigurácií faktorov a ich úrovní, ktoré vyhovujú špecifikáciám návrhu produktu, možno využiť navrhovanie experimentov (DOE – *Design of Experiements*). Základným cieľom aplikácie DOE je nájsť faktory a ich úrovne, ktoré vyhovujú špecifikáciám, s minimálnymi stratami, ktoré sú rastúcou funkciou odchýlok sledovaných ukazovateľov kvality od ich cieľových hodnôt. Takú funkciu nazveme stratovou funkciou.

Keď sa uvažuje niekoľko kritérií návrhu, nemusí byť konfigurácia faktorov a ich úrovní, preferovaná podľa jedného kritéria, preferovaná aj podľa ostatných kritérií. Potom treba určiť nejaké celkové hodnotiace kritérium. Jedným z možných prístupov je metóda funkcie užitočnosti[2], v ktorej sú individuálne stratové funkcie agregované do celkovej stratovej funkcie. Budeme predpokladať, že uvažovaná stratová funkcia je funkcia užitočnosti[3]. To vyžaduje zvoliť typ funkcie užitočnosti a váhy.

Budeme uvažovať len aditívnu funkciu užitočnosti. Všeobecne ide o viacrozmernú (viackriteriálnu) funkciu užitočnosti[4] (*multidimensional* (*multiattribute) utility function*).

1 Navrhovanie experimentov a zlepšovanie procesu

Experiment možno definovať ako ľubovolný zásah do systému s cieľom pozorovať alebo merať účinky tohto zásahu.

---

[1] *podrobnejšie pozri napríklad v Terek – Hrnčiarová (2004)*

[2] *podrobnejšie o metóde funkcie užitočnosti, napr. v Fishburn (1970), Keeney-Raiffa (1976), Winston (1994), Clemen – Reilly (2001)*

[3] *Netradične prijmeme predpoklad, že čím menšia je hodnota funkcie užitočnosti výstupu, tým väčšia je jeho užitočnosť. Výstup* x_1 *je preferovaný pred výstupom* x_2 *(*$x_1 \succ x_2$*) vtedy a len vtedy, keď* $u(x_1) < u(x_2)$*, kde u je funkcia užitočnosti.*

[4] *Podrobnejšie o viackriteriálnych funkciách užitočnosti, pozri napríklad v Keeney-Raiffa (1976), Winston (1994), Clemen – Reilly (2001)*

V experimente vystupuje jedna (alebo viac) premenná, ktorá sa nazýva *ozva (ozvová premenná, výstupná premenná)* a reprezentuje výstup z experimentu a niekoľko predikčných premenných.

Predikčná premenná, ktorá sa mení s cieľom posúdiť jej vplyv na ozvu sa nazýva *faktor*. Potenciálne nastavenie, hodnota alebo vymedzenie faktora sa nazýva *úroveň faktora.*

Experiment možno obyčajne rozdeliť na relatívne samostatné časti, ktoré sa nazývajú *pokusy.* V jednom pokuse sú úrovne faktorov konštantné. Pokus možno viac krát opakovať.

2 Viacfaktorové experimenty a robustný návrh

Keď nás zaujíma vplyv súčasne viacerých faktorov na ozvu, možno realizovať viacfaktorové experimenty.

Termínom *šum* označuje G. Taguchi každú príčinu variability výkonovej charakteristiky produktu v rukách používateľa bez ohľadu na to, či má pôvod vo výrobnom procese, v spôsobe používania produktu používateľom, alebo v vplyve okolia.

Robustným nazýva Taguchi každý produkt alebo proces, ktorý dobre funguje (variabilita jeho výkonových charakteristík je malá) v podmienkach prítomnosti šumu.

Vhodným nástrojom na získanie robustného návrhu produktu alebo procesu je navrhovanie experimentov.

Metóda využitia DOE pri robustných návrhoch[5] sa niekedy nazýva Taguchiho metóda.

Postup viackriteriálneho rozhodovania v Taguchiho metóde sa pokúsime vysvetliť na jednoduchom príklade.

Príklad. Uvažujeme 4 faktory: F_1, F_2, F_3, F_4 a dve ozvy K_1, K_2 – výkonové charakteristiky typu „the smaller the better“[6], ktoré považujeme za kritériá rozhodovania. Experiment sa realizoval podľa Taguchiho ortogonálneho návrhu $L_9\left(3^4\right)$. V tabuľke 1 je uvedený návrh $L_9\left(3^4\right)$ (namiesto číslic 1, 2, 3 sme na označenie úrovní faktorov použili číslice (– 1), 0 a 1[7]) a výsledky experimentu – hodnoty kritérií – oziev K_1 *a* K_2, v posledných dvoch stĺpcoch

Tabuľka 1. *Návrh a výsledky experimentu*

Pokus číslo	F_1	F_2	F_3	F_4	K_1	K_2
1	– 1	–1	– 1	– 1	350,81	22,14
2	– 1	0	0	0	353,53	26,69
3	– 1	1	1	1	346,05	22,19
4	0	–1	0	1	348,51	26,63
5	0	0	1	–1	348,70	25,77
6	0	1	–1	0	351,57	25,89
7	1	–1	1	0	354,33	17,59
8	1	0	– 1	1	345,21	17,77
9	1	1	0	–1	350,00	18,59

[5] *V tejto súvislosti môžu byť zaujímavé aj tolerančné intervaly. Podrobnejšie napríklad v Garaj – Janiga (2004) a Garaj – Janiga (2005).*

[6] *Možno uvažovať aj kritériá typu „the larger the better“ a „nominal is best“.*

[7] *pozri napr. v Grove – Davis (1992)*

3 Vickriteriálne rozhodovanie pomocou funkcie užitočnosti

Problém viackriteriálneho rozhodovania možno transformovať na problém rozhodovania s jediným kritériom. V príklade by sme riešili takúto úlohu

$$\begin{aligned} &u\{K_1(F_1,F_2,F_3,F_4),K_2(F_1,F_2,F_3,F_4)\}\rightarrow \min \\ &F_1\in(-1,0,1) \\ &F_2\in(-1,0,1) \\ &F_3\in(-1,0,1) \\ &F_4\in(-1,0,1) \end{aligned} \qquad (1)$$

kde *u* je funkcia užitočnosti.

Viacrozmernú funkciu užitočnosti[8] môžeme definovať ako aditívnu[9]:

$$u_a = w_1K_1 + w_2K_2 \qquad (2)$$

kde w_1, w_2 sú váhy kritérií.

Ide o najjednoduchšiu formu funkcie užitočnosti. Váhy možno interpretovať ako číselné ohodnotenie relatívnej dôležitosti kritéria pre rozhodovateľa[10]. Obyčajne sa volia tak, že sú to kladné čísla, ktorých súčet sa rovná 1.

Keď rozhodovateľ nie je presvedčený o vhodnosti lineárnej formy funkcie užitočnosti vo „svojej“ aplikácii (napríklad keď nemožno prijať predpoklad o aditívnej nezávislosti kritérií), odporúča sa vyskúšať aj iné jej formy[11].

4 Aplikácia aditívnej funkcie užitočnosti

V súlade s Taguchiho koncepciou nebudeme uvažovať interakcie. Pre každý faktor separovane odhadneme pre každú úroveň veľkosť straty[12], pomocou hodnoty výberového priemeru.

Výsledky pre „náš“ príklad sú v tabuľke 2. Výraznejším písmom sú v tabuľke 2 zapísané minimálne hodnoty, pre kombinácie kritérium – úroveň faktora. V tabuľke 2 vidieť, že ani jeden faktor nemá optimálnu tú istú úroveň pre jedno aj druhé kritérium.

---

[8] *Predpokladá vzájomnú nezávislosť kritérií podľa užitočnosti (mutually utility independence). Keď podmienená funkcia užitočnosti nezávisí od iného kritéria, ide o nezávislosť podľa užitočnosti. Podrobnejšie, napr. v Winston (1994), Clemen-Reilly (2001)*

[9] *Predpokladá aditívnu nezávislosť(additive independence) kritérií, čo je prísnejšia podmienka ako nezávislosť podľa užitočnosti. Aditívna nezávislosť implikuje aditívnu viackriteriálnu funkciu užitočnosti. Aditívna nezávislosť znamená, že preferencie medzi lotériami závisia len od ich marginálnych rozdelení pravdepodobnosti, nie od združeného rozdelenia pravdepodobnosti viacrozmerného výstupu Podrobnejšie, napr. v Winston (1994), Clemen-Reilly (2001)*

[10] *Možno ich určiť napríklad metódou rovnakého oceňovania (pricing out) alebo metódou výkyvného váženia (swing-weighting).*

[11] *Multiplikatívnu alebo typu minimax. Podrobnejšie v Milani – El-Lahham – Nemes (2004).*

[12] *Čím väčšia je hodnota kritéria, tým väčšia je hodnota straty, pretože cieľová hodnota každého kritéria je 0.*

Tabuľka 2. *Straty pre kombinácie kritérií a úrovní faktorov*

	K_1				K_2			
faktory úrovne	F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4
– 1	350,13	351,22	349,20	349,84	23,67	22,12	21,93	22,17
0	349,59	349,15	350,68	353,14	26,10	23,41	23,97	23,39
1	349,85	349,21	349,69	346,59	17,98	22,22	21,85	22,20

Pre kritérium K_1 je optimálna kombinácia úrovní faktorov

$(F_1, F_2, F_3, F_4)^0 = (0, 0, -1, 1)$

a pre kritérium K_2 je optimálna kombinácia úrovní faktorov

$(F_1, F_2, F_3, F_4)^0 = (1, -1, 1, -1)$

Je nevyhnutné hľadať kompromisné riešenie.

Pri aplikácii aditívnej funkcie užitočnosti sa v prvom kroku normujú[13] hodnoty oboch kritérií pre jednotlivé pokusy podľa vzťahu

$$NK_{il} = \frac{K_{il}}{\sum_{l=1}^{9} K_{il}} \qquad \text{pre } i = 1, 2 \ ; \ l = 1, 2, ..., 9 \qquad (3)$$

kde *i* je index kritéria a *l* je index pokusu.

Potom možno hodnoty aditívnej funkcie užitočnosti pre jednotlivé pokusy vypočítať podľa vzťahu

$$Nu_l = w_1 NK_{1l} + w_2 NK_{2l} \qquad \text{pre } l = 1, 2, ..., 9$$

V tabuľke 3 sú pre jednotlivé pokusy normované hodnoty kritérií a hodnoty aditívnej funkcie užitočnosti pre dve kombinácie hodnôt váh.

Potom sa pre každý faktor vypočíta priemerná hodnota aditívnej funkcie užitočnosti pre každú úroveň faktora a nájdu sa optimálne úrovne faktorov. Výsledky výpočtov sú v tabuľke 4.

V tabuľke 4 sú výraznejším písmom uvedené minimálne hodnoty priemerov aditívnej funkcie užitočnosti v jednotlivých stĺpcoch, ktoré určujú optimálne úrovne faktorov.

Pre váhy $w_1 = 0{,}8$, $w_2 = 0{,}2$ je optimálna kombinácia úrovní faktorov

$(F_1, F_2, F_3, F_4)^0 = (1, 1, -1, 1)$

a pre váhy $w_1 = 0{,}6$, $w_2 = 0{,}4$ je optimálna kombinácia úrovní faktorov

[13] *Aby boli hodnoty oboch kritérií merané v stupnici s rovnakým rozsahom.*

$(F_1, F_2, F_3, F_4)^0$ = (1, 1, 1, 1)

Tabuľka 3. *Normované hodnoty kritérií a hodnoty aditívnej funkcie užitočnosti*

pokus číslo	NK_{1l}	NK_{2l}	Nu_l pre $w_1 = 0{,}8$, $w_2 = 0{,}2$	Nu_l pre $w_1 = 0{,}6$, $w_2 = 0{,}4$
1	0,1114	0,1089	0,1109	0,1104
2	0,1123	0,1313	0,1161	0,1199
3	0,1099	0,1092	0,1098	0,1096
4	0,1107	0,1310	0,1148	0,1188
5	0,1107	0,1268	0,1139	0,1171
6	0,1117	0,1274	0,1148	0,1180
7	0,1125	0,0865	0,1073	0,1021
8	0,1096	0,0874	0,1052	0,1007
9	0,1112	0,0915	0,1073	0,1033

Tabuľka 4. *Hodnoty priemerov aditívnej funkcie užitočnosti pre kombinácie úrovní faktorov a váh*

pre váhy	$w_1 = 0{,}8$, $w_2 = 0{,}2$				$w_1 = 0{,}6$, $w_2 = 0{,}4$			
faktory úrovne	F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4
– 1	0,1123	0,1110	**0,1103**	0,1107	0,1133	0,1104	0,1100	0,1103
0	0,1145	0,1117	0,1127	0,1127	0,1180	0,1126	0,1140	0,1133
1	**0,1066**	**0,1106**	0,11033	**0,1099**	**0,1021**	**0,1103**	**0,1096**	**0,1097**

Poďakovanie

Tento článok vznikol s prispením grantovej agentúry VEGA v rámci projektu číslo 1/1247/04 Progresívne štatistické techniky a rozhodovanie v procese zlepšovania kvality a projektu číslo 1/3182/06 Zlepšovanie kvality produkcie strojárskych výrobkov pomocou štatistických metód.

Literatúra

[1] CLEMEN, R. T.: *Making Hard Decisions: An Introduction to Decision Analysis*. Boston: PWS Kent 1991.

[2] CLEMEN, R. T. – REILLY, T.: *Making hard decisions with decision tools*. Duxbury Thomson Learning 2001. ISBN 0-534-42199-7.

[3] FISHBURN, P. C.: *Utility Theory for Decision Making*. New York: Wiley and Sons, 1970.

[4] GARAJ, I. – JANIGA, I.: *Dvojstranné tolerančné medze normálnych rozdelení s neznámymi strednými hodnotami a s neznámym spoločným rozptylom.* Bratislava: STU 2004. ISBN 80-227-2019-4.

[5] GARAJ, I. – JANIGA, I.: *Jednostranné tolerančné medze normálneho rozdelenia s neznámou strednou hodnotou a rozptylom. One sided tolerance limits of normal distribution with unknown mean and variability.* Bratislava: STU 2005. ISBN 80-227-2218-9.

[6] GROVE, D. M. – DAVIS, T. P.: *Engineering, Quality and Experimental Design*. London: Longman Group UK Ltd 1992.

[7] KEENEY, R. L. – RAIFFA, H.: *Decisions with Multiple Objectives.* New York: Wiley and Sons 1976.

[8] LOGOTHETIS, N. – WYNN, H. P.: *Quality through Design*. Oxford: Oxford University Press, 1989.

[9] MILANI, A. S. – EL-LAHHAM, C. – NEMES, J. A.: *Different Approaches in Multiple-Criteria Optimization using the Taguchi Method: A Case Study in a Cold Heading Process*. Journal of Advanced Manufacturing Systems, Vol. 3, No 1 (2004), 53 – 68.

[10] TEREK, M.: *Úvod do analýzy rozhodovania a bayesovskej indukcie.* Bratislava: Ekonóm 2000. ISBN 80-225-1274-5.

[11] TEREK, M. – HRNČIAROVÁ, Ľ.: *Štatistické riadenie kvality*. Bratislava: IURA EDITION 2004. ISBN 80-89047-97-1.

[12] TEREK, M.: *Decision Trees and Influence diagrams in Decision analysis.* Journal of the Applied Mathematics, Statistics and Informatics, Vol. 1, Number 1, pp. 121 – 135.

[13] WHEELER, D.: *Tables of Screening Designs*. Knoxville, Tennessee: SPC Press 1989.

[14] WINSTON, W. L.: *Operations Research. Applications and Algorithms.* Belmont, California: Duxbury Press. 1994. ISBN 0-534-20971-8.

doc. Ing. Milan Terek, PhD., Katedra štatistiky FHI EU v Bratislave, Dolnozemská cesta 1, 852 35 Bratislava, tel.: 02-67295713, e-mail: terek@dec.euba.sk

JEDNOSTRANNÉ PREBERACIE PLÁNY MERANÍM

Ivan Garaj

Abstrakt. In the paper is derived estimate formula for computing the sample size n and the decision number k for samlping inspection by variables for given one-sided tolerance limits. We assume that the parametres of the acceptance sampling plan (P_1 - acceptable quality level; P_2 - limiting quality level; α - producer´s risk; β - consumer´s risk) are known.

Key words: statistical quality control; sampling plan by variables; sample size; decision number; operating characteristic curve.

1. Úvod. Kontrola kvality je činnosť, ktorá zisťuje mieru, resp. stupeň kvalitatívnych vlastností výrobkov, porovnáva zistené výsledky s dopredu určenými požiadavkami a podľa tohto porovnania rozhoduje, či daný výrobok je alebo nie je vhodný na daný účel. Súhrn úkonov, ktorých cieľom je zistiť, či daný výrobok odpovedá dohodnutým vlastnostiam sa nazýva preberacia kontrola (prebierka). Podľa rozsahu kontroly sa potom preberacia kontrola delí na stopercentnú a výberovú. Stopercentnou kontrolou sa rozumie kontrola celého súboru výrobkov a charakterizuje ju veľký objem kontrolných operácií a v súhrne vysoké náklady na technickú kontrolu. Na rozdiel od stopercentnej kontroly sa výberovou kontrolou preveruje iba určitý počet jednotiek vybraných z kontrolovaného súboru dohodnutým spôsobom na základe matematicko-štatistických metód. Zvykne sa tiež nazývať štatistickým riadením kvality.

2. Štatistické riadenie kvality. Štatistické riadenie kvality je časť operatívneho riadenia kvality (operatívne metódy a činnosti používané na splnenie požiadavok na kvalitu), pri ktorej sa používajú postupy matematickej štatistiky. Tieto postupy zahŕňajú použitie rozdelení početností, mier polohy a rozptýlenia, regulačných diagramov, výberových prebierok, korelačnej a regresnej analýzy, analýzy rozptylu, testov významnosti, štatistických tolerančných intervalov, plánovanie experimentov, preberacích plánov atď.

3. Štatistická prebierka. Štatistická prebierka je výberová kontrola, pri ktorej sa urobí rozhodnutie, či prijať alebo neprijať dávku (alebo iné zoskupenie výrobkov, materiálov alebo služieb) na základe výsledkov z výberu alebo výberov odobratých z tejto dávky. Náhodný výber n sa volí podľa vopred určených pravidiel, aby bol dostatočne reprezentatívny. Tieto postupy odstraňujú subjektívne rozhodovania, ktoré sú obyčajne závislé na odhade a skúsenosti kontrolóra. Majú praktický význam predovšetkým pre zabezpečenie vzájomne výhodných vzťahov medzi dodávateľom a odberateľom. Ak výberová kontrola medzi dodávateľom a odberateľom nie je vyslovene dohodnutá vrátane preberacieho plánu, nemá odberateľ právo reklamovať dodávku len na základe náhodnej kontroly. Musí obyčajne vytriediť chybné kusy (napríklad v chemických výrobách to zväčša ani nie je možné) z celej dodávky a až potom reklamovať. Tým vznikajú časové straty, plytvanie pracovnou kapacitou a naviac aj stopercentná kontrola [5] vzhľadom na svoju jednotvárnosť má účinnosť iba na 70-80 %.

4. Ekonomický význam štatistickej prebierky. Ekonomický význam štatistickej prebierky pred inými druhmi kontroly možno zhrnúť do nasledujúcich siedmich bodov [1], [2], [3], [15], [16] a [17].

1. Na základe výberovej kontroly umožňuje objektívne posúdiť, či preberaná dávka výrobkov vyhovuje dohodnutým kvalitatívnym požiadavkám.

2. Zaručuje s dopredu danou pravdepodobnosťou, že budú prijaté iba tie dávky, ktoré spĺňajú požiadavky dohodnuté dopredu medzi dodávateľom a odberateľom.

3. Zaisťuje najhospodárnejšie rozsahy kontroly výrobkov.

4. Účinnosť prebierok nekolíše od jedného prípadu ku druhému, ale je pevná a dopredu daná.

5. Dodávateľovi umožňuje získať prehľad o kvalite dodaných výrobkov, reklamáciách, a. p.

6. Výsledky preberacích plánov ponechávajú zodpovednosť za kvalitu na dodávateľovi, kde správne patrí. Ak nie sú jeho výrobky vyhovujúcej kvality, vzniknú mu ďalšie náklady so zamietnutými dodávkami u odberateľa.

7. Dôsledná aplikácia preberacích plánov znamená pre odberateľa nižšie kontrolné náklady a súčasne možnosť zvýšenia akosti výrobkov u dodávateľa.

Z horeuvedeného vyplýva, že preberacie plány sú metódou optimálneho rozhodovania v zmysle maximálnej hospodárnosti a poskytovaných záruk [1]. Jej bezprostredným cieľom je získať na základe malého počtu kontrol dostatočný prehľad o akosti všetkých vyrábaných alebo dodávaných výrobkov v danom súbore [4]. Ich aplikácia je vhodná tam, kde sa vyrába veľké množstvo rovnakých výrobkov (sériová a hromadná výroba) a kde stopercentná kontrola by bola veľmi prácna a vzhľadom na stálosť výroby a veľký stupeň opakovateľnosti vo výrobe aj málo efektívna. Tejto požiadavke vyhovujú tie výroby, kde sa prevádzajú laboratórne skúšky, deštrukčné skúšky a skúšky so zložitou nákladovou kontrolou.

Základný spôsob delenia štatistických prebierok je podľa spôsobu uskutočňovania kontroly, t. j. podľa toho, či sledovaný znak kvality dodávky výrobkov je nemerateľný (kvalitatívny), alebo merateľný (kvantitatívny). Potom hovoríme o štatistickej prebierke pri kontrole porovnávaním [2] alebo štatistickej prebierky meraním [3]. Pri prebierkach porovnávaním sa výrobky triedia na dobré a chybné a rozhodnutie o prijatí alebo zamietnutí preberaného súboru sa robí na základe celkového počtu zistených chybných výrobkov v kontrolovanom súbore. V nasledujúcom texte sa budeme zaoberať iba štatistickými prebierkami meraním.

5. Štatistické prebierky meraním. Pri štatistickej prebierke meraním sa sledovaný znak kvality výrobku meria a o prijatí alebo zamietnutí preberanej dávky sa rozhoduje na základe hodnôt znaku kvality, ktorý bol vypočítaný z nameraných hodnôt. Tieto prebierky sa obyčajne rozlišujú z nasledujúcich hľadísk:

a) podľa postupu pri zamietnutí preberanej dávky na prebierky opravné a bezopravné;

b) podľa predpísaných náhodných výberov na prebierky jedným, dvojitým, viacnásobným a postupným výberom;

c) vzhľadom k informáciám o rozptyle hodnôt sledovaného kvalitatívneho znaku v preberanej dávke na prebierky pri známom a neznámom rozptyle;

d) vzhľadom k chybám vo výsledkoch analýz skúšobných vzoriek na prebierky, kde uvedené chyby vo výsledkoch sú zanedbateľné a na prebierky, kde tieto chyby nie je možné zanedbať;

e) vzhľadom k predpisu tolerančných medzí pre sledovaný kvalitatívny znak na prebierky pri predpise jednostranných tolerančných medzí (maximálna hodnota USL alebo minimálna hodnota LSL sledovaného kvalitatívneho znaku) a na prebierky pri predpise obojstranných tolerančných medzí (USL a LSL sú predpísané súčasne).

6. Preberací plán. Pravidlo, podľa ktorého sa hodnotí kvalita dávky výrobkov na základe výberovej kontroly sa nazýva preberacím plánom. U každého preberacieho plánu treba stanoviť rozsah náhodného výberu n preberaciu konštantu k (rozhodné číslo), ktoré tvoria preberací plán. Účinnosť každého preberacieho plánu vyjadruje krivka $L(p, n, k)$, ktorú nazývame operatívnou charakteristikou. Udáva pravdepodobnosť, že dávka výrobkov obsahujúca P percent nekvalitných výrobkov bude prijatá pri použití daného preberacieho plánu. Norma STN ISO 3534-2 rozoznáva tri typy operatívnych charakteristík [8]: Typ A je krivka, ktorá udáva pre daný preberací plán pravdepodobnosť, že je splnené preberacie kritérium ako funkcia úrovne kvality dávky. Typ B je krivka, ktorá pre daný preberací plán udáva pravdepodobnosť prijatia dávky ako funkciu úrovne kvality procesu z ktorého dávka pochádza (u niektorých plánov je to krivka udávajúca percento dávok alebo jednotiek výrobku, o ktorých je možné predpokladať, že budú prijaté ako funkcia úrovne kvality procesu). Typ C je krivka, ktorá pre preberacie plány výberovej kontroly pre plynulú výrobu udáva dlhodobé percento výrobkov prijatých v priebehu výrobnej fázy ako funkcie úrovne kvality procesu.

7. Konštrukcia preberacieho plánu. Preberací plán je konštruovaný tak, aby pri kontrole väčšieho počtu preberaných súborov, priemerný počet prípadov, kedy na základe zistenej hodnoty výberového ukazovateľa kvality je dodávateľovi (výrobcovi) vrátená preberaná dávka s $p = P_1$ percentami nezhodných jednotiek, alebo keď na základe zistenej hodnoty výberového ukazovateľa kvality odberateľ príjme dávku s $p = P_2$ nezhodných jednotiek, odpovedal dopredu danej dohode. V prvom prípade sa jedná o riziko dodávateľa (výrobcu), ktoré označujeme α, v druhom prípade ide o riziko odberateľa, ktoré označujeme β. P_1 nazývame prípustnou úrovňou kvality v preberanej dávke, ktorú odberateľ príjme s vysokou pravdepodobnosťou $1-\alpha$ a P_2 hraničnou (neprípustnou) úrovňou kvality v preberanej dávke, ktorú odberateľ príjme najviac s malou pravdepodobnosťou β. Snahou výrobcu a odberateľa je správnou voľbou P_1 zabrániť zamietnutiu dobrých dávok odberateľom. Vyššie P_2 určuje odberateľ z hľadiska ďalšieho použitia výrobkov, resp. suroviny. Preberacie plány, v ktorých dopredu zadávame všetky štyri parametre α, β, P_1 *a* P_2 sa nazývajú (P_1, P_2), pričom. rizika α a β sa v praxi volia $\leq 0{,}10$.

8. Preberacie plány (P_1, P_2). Preberacie plány, v ktorých dopredu zadávame všetky štyri parametre α, β, P_1 *a* P_2 sa nazývajú (P_1, P_2). Rizika α a β sa v praxi volia $\leq 0{,}10$. Žiadame pritom, aby platili vzťahy

$$L(P_1, n, k) = 1 - \alpha \tag{1}$$

$$L(P_2, n, k) = \beta \tag{2}$$

Uvažujme prípad, keď ukazovateľ kvality (náhodná premenná X) má v dávke normálne rozdelenie $N(\mu, \sigma^2)$, kde parametre μ a σ^2 sú neznáme a je predpísaná horná tolerančná medza USL (**U**pper **S**pecification **L**imit - obyčajne daná normou) a výsledky analýz (meraní) nie sú zaťažené chybami, alebo chyby, ktorými sú tieto analýzy zaťažené sú zanedbateľné. Potom podiel nezhodných jednotiek v dávke je

$$p = P(X > USL) = 1 - \Phi\left(\frac{USL - \mu}{\sigma}\right) = 1 - \Phi\left(K_{1-p}\right) \tag{3}$$

kde Φ je normovaná distribučná funkcia $N(0, 1)$ a

$$K_{1-p} = \frac{USL - \mu}{\sigma} \tag{4}$$

je kvantil rádu $1 - p$ tohto rozdelenia tabelovaný v [9]. Vzhľadom k (4) možno požiadavku kvality dávky formulovať tak, že budeme prijímať všetky dávky, ktoré spĺňajú podmienku $\mu + \sigma\ K_{1-p} \leq USL$, kde neznáme parametre μ a σ odhadneme pomocou výberových charakteristík

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{a} \quad s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{5}$$

Budeme prijímať všetky dávky, pri ktorých výberovou kontrolou zistíme, že

$$\bar{x} + k\ s \leq USL \tag{6}$$

kde k je rozhodné číslo preberacieho plánu (P_1, P_2), ktoré sa vypočíta tak, aby podmienka (6) spĺňala podmienky (1) a (2) tohto preberacieho plánu. Presný výpočet k a n súvisí s výpočtom kvantilov [10], [14] necentrálneho t-rozdelenia, ktorý je veľmi komplikovaný a časovo náročný. Jednoduché, no pre malé n dosť nepresné vzťahy boli získané [6] ,[7] aproximáciou štatistiky $\bar{x} + k\ s$ normálnym rozdelením $N\left(\mu + k\sigma,\ \frac{\sigma^2}{n} + \frac{k^2\ \sigma^2}{2(n-1)}\right)$

$$k \approx \frac{K_{1-\alpha}\ K_{1-P_2} + K_{1-\beta}\ K_{1-P_1}}{K_{1-\alpha} + K_{1-\beta}} \tag{7}$$

$$n \approx \left(1 + \frac{k^2}{2}\right) \left(\frac{K_{1-\alpha} + K_{1-\beta}}{K_{1-P_1} - K_{1-P_2}}\right)^2 \qquad (8)$$

Trochu lepšie výsledky dáva postup uvedený v [7], kde boli použité nevychýlené odhady strednej hodnoty a disperzie štatistiky $\bar{x} + k\,s$.

9. Príklad. Ukazovateľom kvality polyesteru 104 je viskozita [7], meraná v $mPa \cdot s$ (mili Pascal·sekunda) pri teplote 25 °C. Požiadavka normy je, aby horná medza viskozity bola $USL = 1\,000\ mPa \cdot s$. Polyester 104 sa dodáva konštantným počtom sudov, ktorých obsah predstavuje jednu výrobnú dávku (operáciu). Výsledky analýz prakticky nie sú zaťažené chybami. Po dohode dodávateľa s odberateľom boli dohodnuté nasledujúce požiadavky na kvalitu: $\alpha = 0{,}05$; $\beta = 0{,}05$; $P_1 = 0{,}01$; $P_2 = 0{,}04$. Treba navrhnúť preberací plán.

10. Riešenie príkladu. V tabuľkách [9] vyhľadáme $K_{0,95} \approx 1{,}64485$; $K_{0,99} \approx 2{,}32635$ a $K_{0,96} \approx 1{,}75069$. Po dosadení do (7) a (8) dostaneme $k \approx 2{,}0385$ a $n \approx 100{,}51$. Preberací plán je nasledovný: z dávky polyesteru 104 treba odobrať 101 sudov, v každom zmerať viskozitu a z nameraných hodnôt podľa (5) vypočítať výberový aritmetický priemer $\bar{x}$ a výberovú smerodajnú odchýlku s. Potom podľa (6) ak $\bar{x} + k\,s \le 1\,000\ mPa \cdot s$ dávku polyesteru 104 prijímame. V prípade, že (6) neplatí, t. j. $\bar{x} + k\,s > 1\,000\ mPa \cdot s$, dávku polyesteru 104 zamietame. V prípade, že bola predpísaná dolná tolerančná medza LSL (**L**ower **S**pecification **L**imit), dávku sudov polyesteru 104 prijímame, ak $\bar{x} - k\,s \ge LSL$ a zamietame, ak $\bar{x} - k\,s < LSL$.

11. Záver. Ak porovnáme výsledky predchádzajúceho príkladu s presnými výsledkami, ktoré súvisia s výpočtom kvantilov [11], [14] necentrálneho t-rozdelenia, zistíme, že $k = 2{,}0417$ a $n = 102$. Rozhodné číslo k vypočítané podľa vzťahu (7) má dosť veľkú chybu, no rozsah náhodného výberu n vypočítaný podľa vzťahu (8) sa od presného výsledku líši nepatrne.

Poďakovanie

Tento článok vznikol s prispením grantovej agentúry VEGA v rámci projektov číslo 1/1247/04 Progresívne štatistické techniky a rozhodovanie v procese zlepšovania kvality a číslo 1/3182/06 Zlepšovanie kvality produkcie strojárskych výrobkov pomocou štatistických metód.

Literatúra

[1] BOHÁČ, M. Jakost v chemii 1, 1976, s. 11-28.
[2] ČSN 01 0254: Statistická přejímka srovnávaním, Praha. ÚNM, 1976.
[3] ČSN 010256: Zvláštní typy statistických přejímek měřením, Praha. ÚNM, 1969.
[4] LEŠČIŠIN, M. *Organizácia a riadenie výroby v priemyselných podnikoch*. Bratislava. ALFA, 1979.
[5] LINCZÉNYI, A. *Komplexné riadenie kvality*. Bratislava. Práca 1970.

[6] GUENTHER, W. C. Tolerance intervals for univariante distributions. Naval Res. Logist. Quart., 19, 1972, p. 309-333.
[7] GARAJ, I. Preberací plán meraním pri daných jednostranných tolerančných medziach..Statistika, 10/1992, p.431-437.
[8] JANIGA, I., PALENČÁR, R. *Štatistika. Slovník a značky. Časť 2: Štatistické riadenie kvality*. STN ISO 3534-2. Slovenský ústav technickej normalizácie, 1999.
[9] LIKEŠ, J., LAGA, J. *Základní statistické tabulky*. Praha: SNTL, 1978. 488 s. (in Czech).
[10] ODEH, R.E., OWEN, D.B. *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*. New York, Marcel Dekker, 1980.
[11] 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.
[12] 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.
[13] 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, Vyd. STU, 2004, 218 s. ISBN 80-227-2019-4.
[14] GARAJ, I., JANIGA I. *Jednostranné tolerančné medze normálneho rozdelenia s neznámou strednou hodnotou a rozptylom. One Sided Tolerance Limits of Normal Distributions with Unknown Mean and Variability.* Bratislava, Vyd. STU, 2005, 214 s. ISBN 80-227-2218-9.
[15] TEREK, M., HRNČIAROVÁ, Ľ. *Štatistické riadenie kvality*. Vydavateľstvo IURA EDITION, 2004, 234 s. ISBN 80-89047-97-1 (in Slovak).
[16] TEREK, M., HRNČIAROVÁ, Ľ. *Analýza spôsobilosti procesu*. Vydavateľstvo EKONÓM, Ekonomická univerzita v Bratislave, 2001. 205 s. ISBN 80-225-1443-8.
[17] JANIGA I. Metódy štatistickej kontroly akosti. Zborník prednášok zo 4. školy štatistiky EKOMSTAT 91 - Využitie štatistických metód v sociálno-ekonomickej praxi. Trenčianské Teplice, SŠDS 1991.

Kontaktá adresa autora:

RNDr. Ivan Garaj, PhD.,
Ústav informatizácie, automatizácie a matematiky,
FCHPT STU, Radlinského 9, 812 37 Bratislava,
Tel.: +421-2-59325 297, E-mail: ivan.garaj@stuba.sk

POŽIADAVKY NA ROZSAH NÁHODNÉHO VÝBERU JEDNOSTRANNÝCH PREBERACÍCH PLÁNOV MERANÍM

Ivan Garaj

Abstrakt. In the paper is derived an exakt formula for computing the sample size n and the decision number k for samlping inspection by variables. We assume that the parametres of the acceptance sampling plan (P_1-acceptable quality level; P_2- limiting quality level; α - producer´s risk; β - consumer´s risk) are known. Sample size n and decision number k requirements for one-sided sampling plans for $\alpha = \beta = 0{,}05$ are given in table 1.

Key words: sampling plan; operating characteristic curve; one-sided tolerance factors.

1. Úvod. Preberací plán je pravidlo, podľa ktorého sa hodnotí kvalita dávky výrobkov na základe výberovej kontroly. Obsahuje údaje o rozsahu kontroly n (náhodný výber) a rozhodného čísla k. Pri preberacom pláne meraním (P_1, P_2) sa dohodou medzi dodávateľom a odberateľom zadávajú štyri parametre: α - riziko dodávateľa; β - riziko odberateľa; P_1 - prípustná úroveň kvality v preberanej dávke, pričom preberací plán zaručuje dodávateľovi, že jeho dávka výrobkov takejto kvality bude odberateľom zamietnutá s malým rizikom $\alpha \le 0{,}1$ a P_2 - hraničná (neprípustná) úrovňou kvality v preberanej dávke, pričom preberací plán zaručuje odberateľovi, že dávku výrobkov takejto kvality zamietne s vysokou pravdepodobnosťou $1-\beta$ $(\beta \le 0{,}1)$. Podobnou problematikou sa zaoberajú práce [5], [7] a [9].

2. Presný výpočet a . Ak ukazovateľ kvality (náhodná premenná X) má v dávke normálne rozdelenie $N(\mu, \sigma^2)$, pričom parametre μ a σ^2 sú neznáme a je predpísaná horná tolerančná medza USL (**U**pper **S**pecification **L**imit), potom podiel nezhodných jednotiek v dávke je $p = 1 - \Phi(u_{1-p})$, kde u_{1-p} je kvantil rádu $1-p$ normovaného normálneho rozdelenia tabelovaný v [1]. Čísla n a k sa vypočítajú nasledovne: dávku výrobkov chceme prijať vždy vtedy, ak $p \le P_1$ a vždy zamietnuť, ak $p \ge P_2$, kde $0 < P_1 < P_2 < 1$, pričom pre operatívnu charakteristiku $L(p, n, k)$ platia vzťahy

$$1 - \alpha = L(P_1, n, k) = P(\bar{x} + k\, s \le USL \mid p = P_1) \tag{1}$$

$$\beta = L(P_2, n, k) = P(\bar{x} + k\, s \le USL \mid p = P_2) \tag{2}$$

kde parametre μ a σ odhadneme pomocou výberových charakteristík $\bar{x}$ a s. Rozhodne číslo k je ale tiež jednostranný tolerančný činiteľ [2], [3], 4], ktorý [6] vzhľadom k rovniciam (1) a (2) je daný vzťahom

$$k = -\frac{t'_{1-\alpha}\left(n-1; -u_{1-P_1}\sqrt{n}\right)}{\sqrt{n}} = -\frac{t'_{\beta}\left(n-1; -u_{1-P_2}\sqrt{n}\right)}{\sqrt{n}} \tag{3}$$

pričom $t'_{1-\alpha}\left(n-1; -u_{1-P_1}\sqrt{n}\right)$ a $t'_{\beta}\left(n-1; -u_{1-P_2}\sqrt{n}\right)$ sú kvantily [6] necentrálneho t-rozdelenia s parametrom necentrality $\delta = -u_{1-P_1}\sqrt{n}$ resp. $\delta = -u_{1-P_2}\sqrt{n}$.

Riešením rovnice

$$t'_{1-\alpha}\left(n-1;-u_{1-P_1}\sqrt{n}\right) - t'_{\beta}\left(n-1;-u_{1-P_2}\sqrt{n}\right) = 0 \qquad (4)$$

štandartnou numerickou metódou získame n (nie je celé číslo) a potom podľa vzťahu (3) k presnému n vypočítame jednostranný tolerančný činiteľ k.

3. Záver. V tabuľke 1 je uvedený numerický výstup rovnice (4) riešený v systéme Mathematica [8] pre parametre preberacieho plánu $\alpha = \beta = 0{,}05$ a rôzne $P_2 = 1{,}5 \cdot P_1$. Presne vypočítane hodnoty n boli v druhom stĺpci zaokrúhľované nahor. Tretí stĺpec bol vypočítaný pomocou presných hodnôt druhého stĺpca. Výpočet je časovo náročný (pre zaujímavosť na výkonnom počítači pre $n = 500$ jeden riadok trvá aj 15 hodín).

Poďakovanie

Tento článok vznikol s prispením grantovej agentúry VEGA v rámci projektov číslo 1/1247/04 Progresívne štatistické techniky a rozhodovanie v procese zlepšovania kvality a číslo 1/3182/06 Zlepšovanie kvality produkcie strojárskych výrobkov pomocou štatistických metód.

Literatúra

[1] LIKEŠ, J., LAGA, J. *Základní statistické tabulky*. Praha: SNTL, 1978. 488 s. (in Czech).
[2] JÍLEK, M.: *Statistické toleranční meze*. Praha: SNTL, 1988. 275 s. (in Czech).
[3] ODEH, R.E., OWEN, D.B. *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*. New York, Marcel Dekker, 1980.
[4] 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.
[5] 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.
[6] GARAJ, I., JANIGA I. *Jednostranné tolerančné medze normálneho rozdelenia s neznámou strednou hodnotou a rozptylom. One Sided Tolerance Limits of Normal Distributions with Unknown Mean and Variability.* Bratislava, Vyd. STU, 2005, 214 s. ISBN 80-227-2218-9.
[7] 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, Vyd. STU, 2004, 218 s. ISBN 80-227-2019-4.
[8] WOLFRAM, S. *The Mathematica Book*. 3rd ed. Wolfram Media/Cambridge University Press, 1996. 1403 p. ISBN 0-521-58889-8.
[9] TEREK, M., HRNČIAROVÁ, Ľ. *Štatistické riadenie kvality*. Vydavateľstvo IURA EDITION, 2004, 234 s. ISBN 80-89047-97-1 (in Slovak).

Kontaktá adresa autora:
RNDr. Ivan Garaj, PhD.,
Ústav informatizácie, automatizácie a matematiky,
FCHPT STU, Radlinského 9, 812 37 Bratislava,
Tel.: +421-2-59325 297, E-mail: ivan.garaj@stuba.sk

TABUĽKA 1 – DOPORUČENÉ ROZSAHY VÝBEROV A ROZHODNÉ ČÍSLA JEDNOSTRANNÝCH PREBERACÍCH PLÁNOV MERANÍM (P_1, P_2)

$\alpha = 0{,}05; \quad \beta = 0{,}05$

P_2	n	k	P_2	n	k	P_2	n	k	P_2	n	k	P_2	n	k
$P_1 = 0{,}0025$			$P_1 = 0{,}0025$			$P_1 = 0{,}0025$			$P_1 = 0{,}0050$			$P_1 = 0{,}0075$		
0,0037	3110	2,7428	0,0097	213	2,5742	0,1300	12	1,9932	0,0215	131	2,3024	0,0175	370	2,2713
0,0038	2712	2,7384	0,0098	209	2,5723	0,1400	12	1,9718	0,0220	126	2,2977	0,0180	344	2,2657
0,0039	2392	2,7340	0,0099	206	2,5704	0,1500	11	1,9514	0,0225	121	2,2931	0,0185	321	2,2602
0,0040	2131	2,7298	0,0100	202	2,5685	0,1600	10	1,9321	0,0230	117	2,2886	0,0190	301	2,2548
0,0041	1914	2,7256	0,0105	187	2,5595	0,1700	10	1,9135	0,0235	113	2,2841	0,0195	283	2,2495
0,0042	1732	2,7216	0,0110	173	2,5508	0,1800	9	1,8957	0,0240	110	2,2797	0,0200	267	2,2444
0,0043	1578	2,7176	0,0115	162	2,5425	0,1900	9	1,8785	0,0245	106	2,2754	0,0205	252	2,2393
0,0044	1445	2,7137	0,0120	152	2,5345	0,2000	9	1,8619	0,0250	103	2,2712	0,0210	239	2,2344
0,0045	1331	2,7099	0,0125	143	2,5268	0,2500	7	1,7859	0,0255	100	2,2671	0,0215	227	2,2296
0,0046	1231	2,7061	0,0130	135	2,5193	0,3000	6	1,7183	0,0260	97	2,2630	0,0220	216	2,2248
0,0047	1144	2,7025	0,0135	128	2,5121	$P_1 = 0{,}0050$			0,0265	95	2,2590	0,0225	206	2,2202
0,0048	1067	2,6989	0,0140	122	2,5052				0,0270	92	2,2550	0,0230	197	2,2156
0,0049	998	2,6953	0,0145	116	2,4984	0,0075	2176	2,5043	0,0275	90	2,2512	0,0235	188	2,2111
0,0050	937	2,6919	0,0150	111	2,4918	0,0076	2035	2,5019	0,0280	87	2,2473	0,0240	181	2,2068
0,0051	882	2,6885	0,0155	106	2,4855	0,0077	1908	2,4996	0,0285	85	2,2436	0,0245	173	2,2024
0,0052	833	2,6851	0,0160	101	2,4793	0,0078	1793	2,4972	0,0290	83	2,2398	0,0250	167	2,1982
0,0053	788	2,6818	0,0165	97	2,4733	0,0079	1690	2,4949	0,0295	81	2,2362	0,0255	160	2,1940
0,0054	747	2,6786	0,0170	94	2,4674	0,0080	1596	2,4926	0,0300	79	2,2326	0,0260	155	2,1899
0,0055	710	2,6754	0,0175	90	2,4617	0,0081	1511	2,4904	0,0350	64	2,1991	0,0265	149	2,1859
0,0056	676	2,6723	0,0180	87	2,4561	0,0082	1433	2,4882	0,0400	54	2,1694	0,0270	144	2,1819
0,0057	645	2,6692	0,0185	84	2,4506	0,0083	1361	2,4859	0,0450	47	2,1427	0,0275	139	2,1780
0,0058	617	2,6662	0,0190	82	2,4453	0,0084	1296	2,4838	0,0500	42	2,1183	0,0280	135	2,1742
0,0059	590	2,6632	0,0195	79	2,4401	0,0085	1235	2,4816	0,0550	37	2,0958	0,0285	131	2,1704
0,0060	566	2,6603	0,0200	77	2,4350	0,0086	1179	2,4795	0,0600	34	2,0750	0,0290	127	2,1667
0,0061	543	2,6574	0,0205	74	2,4301	0,0087	1128	2,4774	0,0650	31	2,0555	0,0295	123	2,1630
0,0062	522	2,6546	0,0210	72	2,4252	0,0088	1080	2,4753	0,0700	29	2,0372	0,0300	120	2,1594
0,0063	503	2,6517	0,0215	70	2,4204	0,0089	1035	2,4732	0,0750	26	2,0199	0,0305	116	2,1558
0,0064	484	2,6490	0,0220	68	2,4157	0,0090	994	2,4711	0,0800	25	2,0035	0,0310	113	2,1523
0,0065	467	2,6462	0,0225	67	2,4112	0,0091	955	2,4691	0,0850	23	1,9878	0,0315	110	2,1488
0,0066	451	2,6435	0,0230	65	2,4067	0,0092	919	2,4671	0,0900	22	1,9729	0,0320	107	2,1454
0,0067	436	2,6409	0,0235	63	2,4022	0,0093	885	2,4651	0,0950	21	1,9586	0,0325	104	2,1420
0,0068	422	2,6383	0,0240	62	2,3979	0,0094	853	2,4631	0,1000	20	1,9449	0,0330	102	2,1386
0,0069	409	2,6357	0,0245	61	2,3936	0,0095	823	2,4612	0,1100	18	1,9189	0,0335	99	2,1353
0,0070	396	2,6331	0,0250	59	2,3895	0,0096	795	2,4592	0,1200	16	1,8947	0,0340	97	2,1321
0,0071	384	2,6306	0,0255	58	2,3853	0,0097	768	2,4573	0,1300	15	1,8719	0,0345	95	2,1289
0,0072	373	2,6281	0,0260	57	2,3813	0,0098	743	2,4554	0,1400	14	1,8503	0,0350	93	2,1257
0,0073	363	2,6257	0,0265	55	2,3773	0,0099	720	2,4535	0,1500	13	1,8299	0,0355	91	2,1226
0,0074	352	2,6232	0,0270	54	2,3734	0,0100	697	2,4516	0,1600	12	1,8104	0,0360	89	2,1195
0,0075	343	2,6208	0,0275	53	2,3695	0,0105	602	2,4425	0,1700	12	1,7917	0,0365	87	2,1164
0,0076	334	2,6185	0,0280	52	2,3658	0,0110	527	2,4338	0,1800	11	1,7737	0,0370	85	2,1134
0,0077	325	2,6161	0,0285	51	2,3620	0,0115	467	2,4254	0,1900	10	1,7564	0,0375	83	2,1104
0,0078	317	2,6138	0,0290	50	2,3583	0,0120	419	2,4174	0,2000	10	1,7397	0,0380	82	2,1074
0,0079	309	2,6115	0,0295	49	2,3547	0,0125	379	2,4096	0,2500	8	1,6633	0,0385	80	2,1045
0,0080	302	2,6092	0,0300	49	2,3511	0,0130	345	2,4021	0,3000	7	1,5954	0,0390	78	2,1016
0,0081	295	2,6070	0,0350	41	2,3179	0,0135	316	2,3948				0,0395	77	2,0987
0,0082	288	2,6048	0,0400	36	2,2884	0,0140	292	2,3878	$P_1 = 0{,}0075$			0,0400	75	2,0959
0,0083	281	2,6026	0,0450	32	2,2619	0,0145	271	2,3810				0,0405	74	2,0930
0,0084	275	2,6004	0,0500	29	2,2377	0,0150	252	2,3744	0,0110	2034	2,3616	0,0410	73	2,0903
0,0085	269	2,5983	0,0550	27	2,2154	0,0155	236	2,3680	0,0115	1615	2,3532	0,0415	71	2,0875
0,0086	263	2,5962	0,0600	25	2,1947	0,0160	221	2,3617	0,0120	1322	2,3451	0,0420	70	2,0848
0,0087	258	2,5941	0,0650	23	2,1754	0,0165	208	2,3557	0,0125	1108	2,3372	0,0425	69	2,0821
0,0088	253	2,5920	0,0700	21	2,1572	0,0170	197	2,3497	0,0130	947	2,3297	0,0430	68	2,0794
0,0089	247	2,5899	0,0750	20	2,1400	0,0175	187	2,3440	0,0135	821	2,3224	0,0435	67	2,0768
0,0090	243	2,5879	0,0800	19	2,1237	0,0180	177	2,3384	0,0140	722	2,3154	0,0440	66	2,0742
0,0091	238	2,5859	0,0850	18	2,1082	0,0185	169	2,3329	0,0145	641	2,3085	0,0445	65	2,0716
0,0092	233	2,5839	0,0900	17	2,0934	0,0190	161	2,3275	0,0150	575	2,3019	0,0450	64	2,0690
0,0093	229	2,5819	0,0950	16	2,0792	0,0195	154	2,3223	0,0155	520	2,2954	0,0455	63	2,0664
0,0094	225	2,5799	0,1000	16	2,0656	0,0200	147	2,3172	0,0160	473	2,2892	0,0460	62	2,0639
0,0095	221	2,5780	0,1100	14	2,0398	0,0205	141	2,3121	0,0165	434	2,2831	0,0465	61	2,0614
0,0096	217	2,5761	0,1200	13	2,0158	0,0210	136	2,3072	0,0170	400	2,2771	0,0470	60	2,0589

TABUĽKA 1 (pokračovanie)

$\alpha = 0{,}05; \quad \beta = 0{,}05$

P_2	n	k	P_2	n	k	P_2	n	k	P_2	n	k	P_2	n	k
$P_1 = 0{,}0075$			$P_1 = 0{,}0100$			$P_1 = 0{,}0100$			$P_1 = 0{,}0150$			$P_1 = 0{,}0200$		
0,0475	59	2,0565	0,0310	165	2,0983	0,1200	22	1,7644	0,0490	116	1,9150	0,0460	220	1,8708
0,0480	58	2,0540	0,0315	159	2,0948	0,1300	20	1,7414	0,0495	114	1,9126	0,0465	213	1,8682
0,0485	57	2,0516	0,0320	154	2,0914	0,1400	18	1,7197	0,0500	112	1,9102	0,0470	207	1,8657
0,0490	56	2,0492	0,0325	150	2,0880	0,1500	17	1,6990	0,0525	102	1,8986	0,0475	202	1,8632
0,0495	56	2,0468	0,0330	145	2,0846	0,1600	16	1,6793	0,0550	93	1,8874	0,0480	196	1,8607
0,0500	55	2,0445	0,0335	141	2,0813	0,1700	15	1,6605	0,0575	86	1,8767	0,0485	191	1,8583
0,0525	51	2,0330	0,0340	137	2,0781	0,1800	14	1,6424	0,0600	79	1,8663	0,0490	186	1,8558
0,0550	48	2,0219	0,0345	133	2,0748	0,1900	13	1,6249	0,0625	74	1,8562	0,0495	181	1,8534
0,0575	46	2,0112	0,0350	130	2,0716	0,2000	12	1,6081	0,0650	69	1,8465	0,0500	177	1,8510
0,0600	43	2,0009	0,0355	126	2,0685	0,2500	10	1,5310	0,0675	65	1,8371	0,0505	172	1,8487
0,0650	39	1,9814	0,0360	123	2,0654	0,3000	8	1,4625	0,0700	61	1,8279	0,0510	168	1,8463
0,0700	36	1,9629	0,0365	120	2,0623				0,0725	58	1,8190	0,0515	164	1,8440
0,0750	33	1,9456	0,0370	117	2,0593	$P_1 = 0{,}0150$			0,0750	54	1,8103	0,0520	160	1,8417
0,0800	30	1,9291	0,0375	114	2,0562				0,0775	52	1,8019	0,0525	157	1,8394
0,0850	28	1,9134	0,0380	111	2,0533	0,0220	1419	2,0924	0,0800	49	1,7936	0,0530	153	1,8371
0,0900	26	1,8984	0,0385	109	2,0503	0,0230	1125	2,0831	0,0825	47	1,7856	0,0535	150	1,8348
0,0950	25	1,8840	0,0390	106	2,0474	0,0240	919	2,0741	0,0850	45	1,7778	0,0540	146	1,8326
0,1000	23	1,8702	0,0395	104	2,0445	0,0250	769	2,0655	0,0875	43	1,7701	0,0545	143	1,8303
0,1100	21	1,8441	0,0400	102	2,0417	0,0260	656	2,0571	0,0900	41	1,7626	0,0550	140	1,8281
0,1200	19	1,8197	0,0405	100	2,0389	0,0270	568	2,0491	0,0925	39	1,7553	0,0555	137	1,8259
0,1300	18	1,7968	0,0410	97	2,0361	0,0280	498	2,0412	0,0950	38	1,7481	0,0560	135	1,8238
0,1400	16	1,7752	0,0415	95	2,0333	0,0290	442	2,0337	0,0975	37	1,7411	0,0565	132	1,8216
0,1500	15	1,7546	0,0420	94	2,0306	0,0300	396	2,0263	0,1000	35	1,7342	0,0570	129	1,8194
0,1600	14	1,7350	0,0425	92	2,0279	0,0305	376	2,0227	0,1100	31	1,7078	0,0575	127	1,8173
0,1700	13	1,7162	0,0430	90	2,0252	0,0310	357	2,0191	0,1200	27	1,6832	0,0580	124	1,8152
0,1800	12	1,6981	0,0435	88	2,0225	0,0315	340	2,0156	0,1300	25	1,6600	0,0585	122	1,8131
0,1900	12	1,6808	0,0440	86	2,0199	0,0320	325	2,0121	0,1400	22	1,6381	0,0590	120	1,8110
0,2000	11	1,6640	0,0445	85	2,0173	0,0325	310	2,0087	0,1500	20	1,6174	0,0595	118	1,8089
0,2500	9	1,5872	0,0450	83	2,0147	0,0330	297	2,0054	0,1600	19	1,5975	0,0600	115	1,8069
0,3000	7	1,5190	0,0455	82	2,0121	0,0335	285	2,0020	0,1700	17	1,5785	0,0625	106	1,7968
$P_1 = 0{,}0100$			0,0460	80	2,0096	0,0340	273	1,9987	0,1800	16	1,5603	0,0650	97	1,7870
			0,0465	79	2,0071	0,0345	263	1,9955	0,1900	15	1,5427	0,0675	90	1,7775
			0,0470	78	2,0046	0,0350	253	1,9923	0,2000	14	1,5258	0,0700	84	1,7683
0,0150	1565	2,2485	0,0475	76	2,0021	0,0355	243	1,9891	0,2500	11	1,4482	0,0725	78	1,7593
0,0160	1145	2,2357	0,0480	75	1,9997	0,0360	235	1,9860	0,3000	9	1,3793	0,0750	74	1,7506
0,0170	884	2,2236	0,0485	74	1,9973	0,0365	226	1,9829				0,0775	69	1,7422
0,0180	710	2,2122	0,0490	73	1,9948	0,0370	219	1,9798	$P_1 = 0{,}0200$			0,0800	65	1,7339
0,0190	587	2,2012	0,0495	72	1,9925	0,0375	212	1,9768				0,0825	62	1,7258
0,0200	496	2,1908	0,0500	70	1,9901	0,0380	205	1,9738	0,0300	1063	1,9676	0,0850	59	1,7179
0,0205	459	2,1857	0,0525	65	1,9785	0,0385	198	1,9708	0,0310	901	1,9604	0,0875	56	1,7102
0,0210	427	2,1808	0,0550	61	1,9674	0,0390	192	1,9679	0,0320	776	1,9534	0,0900	53	1,7027
0,0215	399	2,1759	0,0575	57	1,9567	0,0395	186	1,9650	0,0330	677	1,9466	0,0925	51	1,6953
0,0220	374	2,1711	0,0600	54	1,9464	0,0400	181	1,9621	0,0340	597	1,9399	0,0950	48	1,6881
0,0225	351	2,1665	0,0625	51	1,9364	0,0405	176	1,9593	0,0350	532	1,9335	0,0975	46	1,6810
0,0230	331	2,1619	0,0650	48	1,9267	0,0410	171	1,9565	0,0360	478	1,9271	0,1000	45	1,6741
0,0235	312	2,1574	0,0675	45	1,9173	0,0415	166	1,9537	0,0370	432	1,9209	0,1100	38	1,6476
0,0240	296	2,1530	0,0700	43	1,9082	0,0420	162	1,9510	0,0380	394	1,9149	0,1200	33	1,6229
0,0245	281	2,1487	0,0725	41	1,8994	0,0425	157	1,9482	0,0390	361	1,9090	0,1300	30	1,5996
0,0250	267	2,1444	0,0750	39	1,8908	0,0430	153	1,9455	0,0400	332	1,9032	0,1400	27	1,5776
0,0255	254	2,1402	0,0775	38	1,8824	0,0435	150	1,9429	0,0405	319	1,9003	0,1500	24	1,5567
0,0260	243	2,1361	0,0800	36	1,8742	0,0440	146	1,9402	0,0410	307	1,8975	0,1600	22	1,5368
0,0265	232	2,1321	0,0825	35	1,8662	0,0445	142	1,9376	0,0415	296	1,8947	0,1700	20	1,5177
0,0270	222	2,1281	0,0850	33	1,8584	0,0450	139	1,9350	0,0420	285	1,8919	0,1800	19	1,4994
0,0275	213	2,1242	0,0875	32	1,8508	0,0455	136	1,9324	0,0425	275	1,8892	0,1900	17	1,4817
0,0280	205	2,1203	0,0900	31	1,8434	0,0460	133	1,9299	0,0430	266	1,8865	0,2000	16	1,4647
0,0285	197	2,1165	0,0925	30	1,8361	0,0465	130	1,9273	0,0435	257	1,8838	0,2500	12	1,3866
0,0290	190	2,1128	0,0950	29	1,8290	0,0470	127	1,9248	0,0440	249	1,8812	0,3000	9	1,3174
0,0295	183	2,1091	0,0975	28	1,8220	0,0475	124	1,9224	0,0445	241	1,8785			
0,0300	176	2,1054	0,1000	27	1,8151	0,0480	121	1,9199	0,0450	234	1,8759			
0,0305	170	2,1019	0,1100	24	1,7889	0,0485	119	1,9174	0,0455	226	1,8733			

TABUĽKA 1 (pokračovanie)

$\alpha = 0{,}05; \quad \beta = 0{,}05$

P_2	n	k	P_2	n	k	P_2	n	k	P_2	n	k	P_2	n	k
$P_1 = 0{,}0250$			$P_1 = 0{,}0250$			$P_1 = 0{,}0300$			$P_1 = 0{,}0400$			$P_1 = 0{,}0400$		
0,0370	993	1,8736	0,0875	72	1,6621	0,0650	198	1,6989	0,0630	523	1,6410	0,0930	130	1,5387
0,0380	863	1,8676	0,0900	68	1,6546	0,0655	193	1,6970	0,0635	503	1,6390	0,0935	128	1,5372
0,0390	759	1,8616	0,0925	64	1,6472	0,0660	189	1,6951	0,0640	485	1,6370	0,0940	126	1,5358
0,0400	674	1,8558	0,0950	61	1,6399	0,0665	185	1,6932	0,0645	468	1,6350	0,0945	124	1,5343
0,0410	603	1,8501	0,0975	58	1,6328	0,0670	181	1,6913	0,0650	452	1,6331	0,0950	122	1,5329
0,0420	544	1,8445	0,1000	56	1,6258	0,0675	177	1,6894	0,0655	437	1,6311	0,0955	121	1,5314
0,0430	494	1,8391	0,1100	47	1,5992	0,0680	173	1,6875	0,0660	422	1,6292	0,0960	119	1,5300
0,0440	451	1,8337	0,1200	40	1,5744	0,0685	170	1,6856	0,0665	409	1,6273	0,0965	117	1,5285
0,0450	414	1,8284	0,1300	35	1,5510	0,0690	167	1,6838	0,0670	396	1,6253	0,0970	116	1,5271
0,0460	382	1,8233	0,1400	31	1,5289	0,0695	163	1,6819	0,0675	384	1,6234	0,0975	114	1,5257
0,0470	354	1,8182	0,1500	28	1,5079	0,0700	160	1,6801	0,0680	372	1,6216	0,0980	113	1,5243
0,0480	329	1,8132	0,1600	25	1,4879	0,0705	157	1,6783	0,0685	361	1,6197	0,0985	111	1,5229
0,0490	307	1,8083	0,1700	23	1,4688	0,0710	154	1,6765	0,0690	350	1,6178	0,0990	110	1,5215
0,0500	287	1,8035	0,1800	21	1,4503	0,0715	151	1,6747	0,0695	340	1,6160	0,0995	108	1,5201
0,0505	278	1,8011	0,1900	20	1,4326	0,0720	148	1,6729	0,0700	331	1,6141	0,1000	107	1,5187
0,0510	270	1,7987	0,2000	18	1,4155	0,0725	146	1,6711	0,0705	322	1,6123	0,1050	94	1,5050
0,0515	261	1,7964	0,2500	13	1,3371	0,0730	143	1,6693	0,0710	313	1,6105	0,1100	84	1,4918
0,0520	254	1,7941	0,3000	10	1,2675	0,0735	140	1,6676	0,0715	304	1,6086	0,1150	76	1,4790
0,0525	246	1,7918				0,0740	138	1,6658	0,0720	296	1,6068	0,1200	69	1,4667
0,0530	239	1,7895	$P_1 = 0{,}0300$			0,0745	136	1,6641	0,0725	289	1,6051	0,1250	63	1,4547
0,0535	233	1,7872	0,0450	819	1,7885	0,0750	133	1,6623	0,0730	282	1,6033	0,1300	58	1,4431
0,0540	227	1,7849	0,0455	774	1,7859	0,0755	131	1,6606	0,0735	274	1,6015	0,1350	53	1,4318
0,0545	220	1,7827	0,0460	732	1,7833	0,0760	129	1,6589	0,0740	268	1,5997	0,1400	49	1,4208
0,0550	215	1,7805	0,0465	693	1,7808	0,0765	127	1,6572	0,0745	261	1,5980	0,1450	46	1,4101
0,0555	209	1,7783	0,0470	658	1,7782	0,0770	125	1,6555	0,0750	255	1,5962	0,1500	43	1,3996
0,0560	204	1,7761	0,0475	625	1,7757	0,0775	123	1,6538	0,0755	249	1,5945	0,1550	40	1,3894
0,0565	199	1,7739	0,0480	596	1,7732	0,0780	121	1,6521	0,0760	243	1,5928	0,1600	38	1,3794
0,0570	194	1,7718	0,0485	569	1,7707	0,0785	119	1,6504	0,0765	238	1,5910	0,1650	36	1,3696
0,0575	189	1,7696	0,0490	543	1,7683	0,0790	117	1,6487	0,0770	233	1,5893	0,1700	34	1,3600
0,0580	185	1,7675	0,0495	520	1,7659	0,0795	115	1,6471	0,0775	227	1,5876	0,1750	32	1,3506
0,0585	181	1,7654	0,0500	498	1,7635	0,0800	113	1,6454	0,0780	223	1,5859	0,1800	30	1,3414
0,0590	177	1,7633	0,0505	477	1,7611	0,0825	105	1,6373	0,0785	218	1,5843	0,1850	29	1,3324
0,0595	173	1,7612	0,0510	458	1,7587	0,0850	98	1,6294	0,0790	214	1,5826	0,1900	27	1,3235
0,0600	169	1,7591	0,0515	440	1,7563	0,0875	92	1,6216	0,0795	209	1,5809	0,1950	26	1,3148
0,0605	165	1,7571	0,0520	423	1,7540	0,0900	86	1,6140	0,0800	204	1,5793	0,2000	25	1,3062
0,0610	162	1,7550	0,0525	408	1,7517	0,0925	81	1,6066	0,0805	200	1,5776	0,2100	23	1,2894
0,0615	158	1,7530	0,0530	393	1,7494	0,0950	77	1,5993	0,0810	196	1,5760	0,2200	21	1,2732
0,0620	155	1,7510	0,0535	379	1,7471	0,0975	73	1,5922	0,0815	193	1,5743	0,2300	20	1,2573
0,0625	152	1,7490	0,0540	366	1,7449	0,1000	69	1,5852	0,0820	189	1,5727	0,2400	18	1,2419
0,0630	149	1,7470	0,0545	353	1,7426	0,1100	57	1,5585	0,0825	185	1,5711	0,2500	17	1,2269
0,0635	146	1,7450	0,0550	342	1,7404	0,1200	48	1,5336	0,0830	182	1,5695	0,2600	16	1,2123
0,0640	143	1,7431	0,0555	331	1,7382	0,1300	42	1,5101	0,0835	178	1,5679	0,2700	15	1,1979
0,0645	140	1,7411	0,0560	320	1,7360	0,1400	36	1,4879	0,0840	175	1,5663	0,2800	14	1,1838
0,0650	138	1,7392	0,0565	310	1,7338	0,1500	32	1,4669	0,0845	172	1,5647	0,2900	14	1,1700
0,0655	135	1,7372	0,0570	301	1,7316	0,1600	29	1,4468	0,0850	169	1,5631	0,3000	13	1,1565
0,0660	133	1,7353	0,0575	292	1,7295	0,1700	26	1,4275	0,0855	166	1,5615			
0,0665	130	1,7334	0,0580	283	1,7273	0,1800	24	1,4091	0,0860	163	1,5600	$P_1 = 0{,}0500$		
0,0670	128	1,7315	0,0585	275	1,7252	0,1900	22	1,3913	0,0865	160	1,5584			
0,0675	126	1,7297	0,0590	268	1,7231	0,2000	20	1,3741	0,0870	157	1,5568	0,0750	563	1,5427
0,0680	124	1,7278	0,0595	260	1,7210	0,2500	15	1,2954	0,0875	155	1,5553	0,0755	544	1,5410
0,0685	121	1,7259	0,0600	253	1,7190	0,3000	11	1,2254	0,0880	152	1,5538	0,0760	525	1,5392
0,0690	119	1,7241	0,0605	246	1,7169				0,0885	150	1,5522	0,0765	508	1,5375
0,0695	117	1,7222	0,0610	240	1,7148	$P_1 = 0{,}0400$			0,0890	147	1,5507	0,0770	491	1,5358
0,0700	115	1,7204	0,0615	234	1,7128				0,0895	145	1,5492	0,0775	476	1,5341
0,0725	107	1,7114	0,0620	228	1,7108	0,0600	668	1,6532	0,0900	142	1,5477	0,0780	461	1,5324
0,0750	99	1,7027	0,0625	222	1,7088	0,0605	640	1,6511	0,0905	140	1,5462	0,0785	447	1,5307
0,0775	92	1,6942	0,0630	217	1,7068	0,0610	613	1,6491	0,0910	138	1,5446	0,0790	433	1,5290
0,0800	86	1,6859	0,0635	212	1,7048	0,0615	588	1,6470	0,0915	136	1,5432	0,0795	420	1,5273
0,0825	81	1,6778	0,0640	207	1,7028	0,0620	565	1,6450	0,0920	134	1,5417	0,0800	408	1,5257
0,0850	76	1,6699	0,0645	202	1,7009	0,0625	543	1,6430	0,0925	132	1,5402	0,0805	397	1,5240

TABUĽKA 1 (dokončenie)

$\alpha = 0{,}05; \quad \beta = 0{,}05$

P_2	n	k	P_2	n	k	P_2	n	k	P_2	n	k	P_2	n	k
$p_1 = 0{,}0500$			$p_1 = 0{,}0500$			$p_1 = 0{,}0750$			$p_1 = 0{,}0800$			$p_1 = 0{,}1000$		
0,0810	385	1,5224	0,2200	26	1,2180	0,2700	25	1,0349	0,2850	24	0,9955	0,1500	302	1,1598
0,0815	375	1,5207	0,2300	24	1,2021	0,2750	24	1,0277	0,2900	23	0,9885	0,1525	276	1,1545
0,0820	365	1,5191	0,2400	22	1,1866	0,2800	23	1,0205	0,2950	22	0,9815	0,1550	254	1,1493
0,0825	355	1,5174	0,2500	20	1,1715	0,2850	22	1,0135	0,3000	21	0,9746	0,1575	235	1,1442
0,0830	346	1,5158	0,2600	19	1,1567	0,2900	22	1,0065	0,3500	15	0,9085	0,1600	217	1,1391
0,0835	337	1,5142	0,2700	18	1,1422	0,2950	21	0,9995	0,4000	12	0,8461	0,1625	202	1,1340
0,0840	328	1,5126	0,2800	16	1,1281	0,3000	20	0,9926				0,1650	188	1,1291
0,0845	320	1,5110	0,2900	16	1,1142	0,3500	15	0,9267	$P_1 = 0{,}0900$			0,1675	176	1,1241
0,0850	312	1,5094	0,3000	15	1,1006	0,4000	11	0,8645				0,1700	165	1,1192
0,0855	305	1,5078	$P_1 = 0{,}0750$			$P_1 = 0{,}0800$			0,1350	336	1,2227	0,1725	155	1,1144
0,0860	298	1,5063							0,1375	304	1,2170	0,1750	146	1,1096
0,0865	291	1,5047	0,1125	399	1,3271	0,1200	376	1,2907	0,1400	277	1,2114	0,1775	138	1,1049
0,0870	284	1,5031	0,1150	355	1,3207	0,1225	337	1,2846	0,1425	254	1,2059	0,1800	130	1,1002
0,0875	277	1,5016	0,1175	319	1,3144	0,1250	304	1,2785	0,1450	234	1,2005	0,1825	123	1,0955
0,0880	271	1,5000	0,1200	288	1,3082	0,1275	276	1,2726	0,1475	216	1,1951	0,1850	117	1,0909
0,0885	265	1,4985	0,1225	262	1,3020	0,1300	252	1,2667	0,1500	200	1,1898	0,1875	111	1,0863
0,0890	259	1,4970	0,1250	239	1,2960	0,1325	231	1,2609	0,1525	186	1,1845	0,1900	106	1,0818
0,0895	254	1,4954	0,1275	219	1,2901	0,1350	213	1,2552	0,1550	174	1,1793	0,1925	101	1,0773
0,0900	249	1,4939	0,1300	202	1,2842	0,1375	197	1,2496	0,1575	163	1,1742	0,1950	97	1,0729
0,0905	243	1,4924	0,1325	187	1,2784	0,1400	183	1,2440	0,1600	153	1,1691	0,1975	92	1,0684
0,0915	233	1,4894	0,1350	174	1,2727	0,1425	170	1,2385	0,1625	144	1,1641	0,2000	88	1,0641
0,0920	229	1,4879	0,1375	162	1,2671	0,1450	159	1,2331	0,1650	135	1,1591	0,2050	81	1,0554
0,0925	224	1,4864	0,1400	151	1,2615	0,1475	149	1,2278	0,1675	128	1,1542	0,2100	75	1,0469
0,0930	220	1,4849	0,1425	142	1,2561	0,1500	140	1,2225	0,1700	121	1,1494	0,2150	69	1,0385
0,0930	220	1,4849	0,1450	133	1,2506	0,1525	132	1,2172	0,1725	115	1,1445	0,2200	65	1,0302
0,0935	216	1,4834	0,1475	126	1,2453	0,1550	124	1,2121	0,1750	109	1,1398	0,2250	60	1,0220
0,0940	212	1,4820	0,1500	119	1,2400	0,1575	118	1,2069	0,1775	104	1,1350	0,2300	56	1,0139
0,0945	208	1,4805	0,1525	112	1,2348	0,1600	111	1,2019	0,1800	99	1,1304	0,2350	53	1,0060
0,0950	204	1,4790	0,1550	106	1,2296	0,1625	106	1,1969	0,1825	94	1,1257	0,2400	50	0,9981
0,0955	200	1,4776	0,1575	101	1,2245	0,1650	100	1,1919	0,1850	90	1,1211	0,2450	47	0,9904
0,0960	196	1,4761	0,1600	96	1,2195	0,1675	96	1,1870	0,1875	86	1,1166	0,2500	44	0,9827
0,0965	193	1,4747	0,1625	92	1,2145	0,1700	91	1,1822	0,1900	83	1,1120	0,2550	42	0,9751
0,0970	189	1,4733	0,1650	87	1,2095	0,1725	87	1,1774	0,1925	79	1,1076	0,2600	40	0,9676
0,0975	186	1,4718	0,1675	83	1,2046	0,1750	83	1,1726	0,1950	76	1,1031	0,2650	38	0,9602
0,0980	183	1,4704	0,1700	80	1,1998	0,1775	80	1,1679	0,1975	73	1,0987	0,2700	36	0,9529
0,0985	180	1,4690	0,1725	76	1,1950	0,1800	76	1,1632	0,2000	70	1,0944	0,2750	34	0,9456
0,0990	177	1,4676	0,1750	73	1,1903	0,1825	73	1,1586	0,2050	65	1,0857	0,2800	33	0,9384
0,0995	174	1,4662	0,1775	70	1,1856	0,1850	70	1,1540	0,2100	61	1,0772	0,2850	31	0,9313
0,1000	171	1,4648	0,1800	68	1,1809	0,1875	68	1,1495	0,2150	57	1,0688	0,2900	30	0,9242
0,1050	146	1,4510	0,1825	65	1,1763	0,1900	65	1,1450	0,2200	53	1,0606	0,2950	29	0,9172
0,1100	127	1,4378	0,1850	63	1,1717	0,1925	63	1,1405	0,2250	50	1,0524	0,3000	27	0,9102
0,1150	111	1,4250	0,1875	60	1,1672	0,1950	61	1,1361	0,2300	47	1,0444	0,3050	26	0,9033
0,1200	99	1,4126	0,1900	58	1,1627	0,1975	59	1,1317	0,2350	44	1,0365	0,3100	25	0,8965
0,1250	89	1,4005	0,1925	56	1,1582	0,2000	57	1,1274	0,2400	42	1,0286	0,3150	24	0,8897
0,1300	80	1,3888	0,1950	54	1,1538	0,2050	53	1,1187	0,2450	40	1,0209	0,3200	23	0,8830
0,1350	73	1,3775	0,1975	53	1,1494	0,2100	50	1,1103	0,2500	38	1,0133	0,3250	23	0,8763
0,1400	67	1,3664	0,2000	51	1,1451	0,2150	47	1,1019	0,2550	36	1,0057	0,3300	22	0,8697
0,1450	61	1,3556	0,2050	48	1,1365	0,2200	44	1,0937	0,2600	34	0,9982	0,3350	21	0,8631
0,1500	57	1,3451	0,2100	45	1,1280	0,2250	42	1,0856	0,2650	33	0,9908	0,3400	20	0,8565
0,1550	52	1,3348	0,2150	42	1,1197	0,2300	39	1,0776	0,2700	31	0,9835	0,3450	20	0,8500
0,1600	49	1,3248	0,2200	40	1,1115	0,2350	37	1,0697	0,2750	30	0,9763	0,3500	19	0,8436
0,1650	46	1,3149	0,2250	38	1,1034	0,2400	35	1,0619	0,2800	28	0,9691	0,3600	18	0,8307
0,1700	43	1,3053	0,2300	36	1,0954	0,2450	34	1,0542	0,2850	27	0,9620	0,3700	17	0,8180
0,1750	40	1,2959	0,2350	34	1,0875	0,2500	32	1,0466	0,2900	26	0,9550	0,3800	16	0,8055
0,1800	38	1,2866	0,2400	33	1,0797	0,2550	31	1,0390	0,2950	25	0,9480	0,3900	15	0,7930
0,1850	36	1,2775	0,2450	31	1,0720	0,2600	29	1,0316	0,3000	24	0,9411	0,4000	14	0,7806
0,1900	34	1,2686	0,2500	30	1,0644	0,2650	28	1,0242	0,3500	17	0,8746	0,4500	11	0,7201
0,1950	32	1,2598	0,2550	28	1,0569	0,2700	27	1,0169	0,4000	13	0,8120	0,5000	9	0,6607
0,2000	31	1,2512	0,2600	27	1,0495	0,2750	26	1,0097	0,4500	10	0,7517			
0,2100	28	1,2343	0,2650	26	1,0421	0,2800	25	1,0026	0,5000	8	0,6927			

Necentralne *t*-rozdelenie v odhadovaní minimálnej detegovateľnej hodnoty pomocou lineárnej kalibračnej funkcie s konštantnou smerodajnou odchýlkou

Ivan Janiga, Ivan Garaj

Abstract. The noncentral *t*-distribution is used in computation of minimum detectable value. The value is obtained from a linear calibration as the smallest value of the net state variable which can be detected with a probability of $1-\beta$ as different from zero. We assume that all measurements of the response variable are independent and normally distributed with constant standard deviation.

1 Úvod

Minimálna detegovateľná hodnota získaná s príslušnej kalibrácie ukazuje schopnosť kalibrovaného meracieho procesu pre príslušné série meraní detegovať hodnotu redukovanej stavovej premennej pozorovaného skutočného stavu odlišného od nuly, teda je to najmenšia hodnota redukovanej stavovej premennej, ktorá môže byť detegovaná s pravdepodobnosťou $1-\beta$ ako hodnota rôzna od nuly (pozri obrázok). V [1] sa predpokladala lineárna kalibračná funkcia s konštantnou smerodajnou odchýlkou $Y_{ij} = a + bx_i + \varepsilon_{ij}$, $i = 1, 2, ..., I$, $j = 1, 2, ..., J$ kde x_i je redukovaná stavová premenná v stave i. Náhodné premenné ε_{ij}, ktoré reprezentujú náhodnú zložku chyby vzorkovania, preparátu a merania, sú nezávislé, normálne rozdelené s nulovou strednou hodnotou a smerodajnou odchýlkou σ, t. j. $\varepsilon_{ij} \sim N(0, \sigma^2)$. Merania ozvovej premennej pre všetky referenčné stavy (I) a preparáty (J) Y_{ij} sú nezávislé náhodné premenné, ktoré majú normálne rozdelenie so strednou hodnotou $E(Y_{ij}) = a + bx_i$ a rozptylom $D(Y_{ij}) = \sigma^2$, ktorý nezávisí od hodnôt x_i, t. j. $Y_{ij} \sim N(a + bx_i, \sigma^2)$, $i = 1, 2, ..., I$, $j = 1, 2, ..., J$.

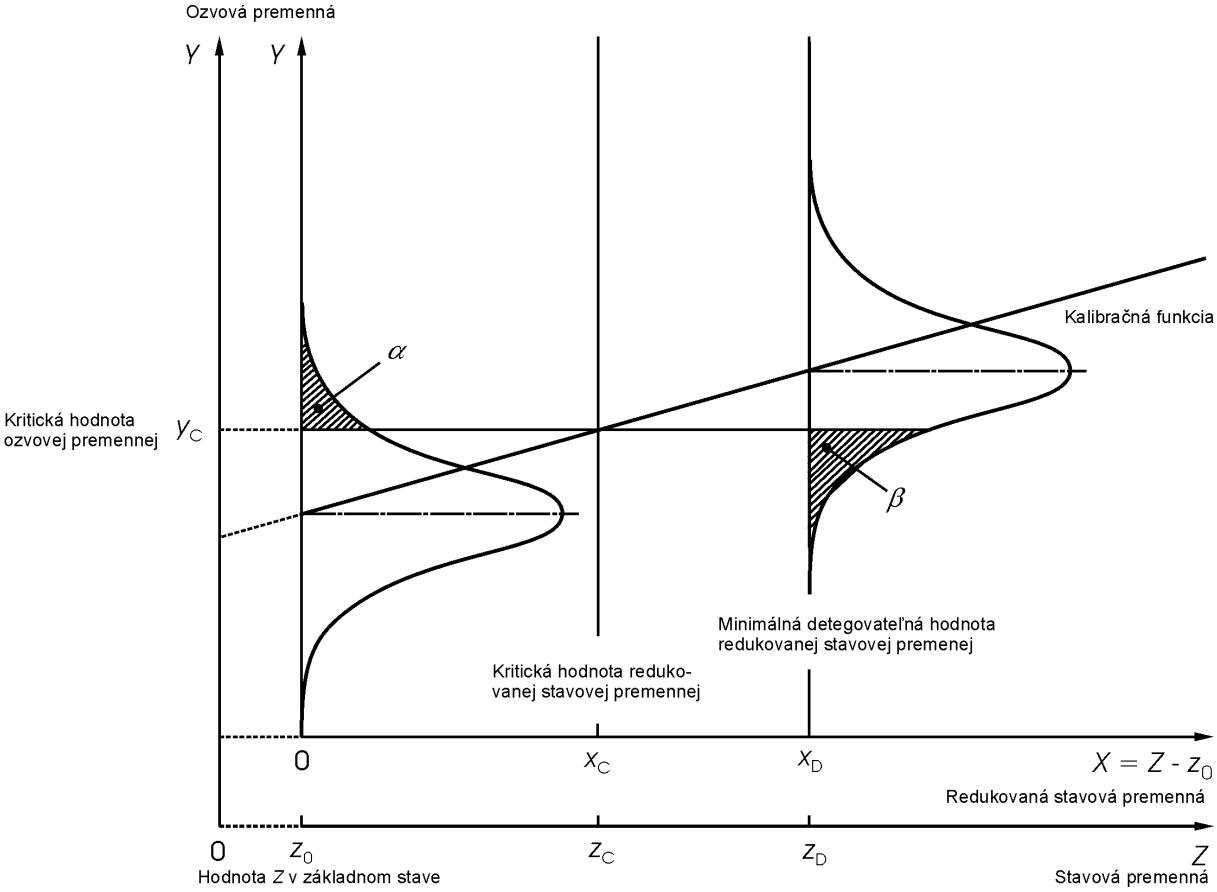
2 Odhad minimálnej detegovateľnej hodnoty z lineárnej kalibračnej funkcie

V tejto časti uvádzame dôležité výsledky z publikácie [1]. Vzájomné vzťahy kritickej hodnoty ozvy y_c, kritickej hodnoty redukovanej stavovej premennej x_c a minimálnej detegovateľnej hodnoty x_d možno vidieť na obrázku 1. Odhad minimálnej detegovateľnej hodnoty sme dostali z odhadu kalibračnej funkcie (pozri obr. 1).

Pre odhady *a*, *b* a σ^2 platí:

$$\hat{b} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}(x_i - \bar{x})(\bar{y}_{ij} - \bar{y})}{s_{xx}}, \quad \hat{a} = \bar{y} - \hat{b}\bar{x}, \quad \hat{\sigma}^2 = \frac{1}{I \cdot J - 2}\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{ij} - \hat{a} - \hat{b}x_i)^2$$

kde $s_{xx} = J\sum_{i=1}^{I}(x_i - \bar{x})^2$.



Obrázok 1

Kritickú hodnotu ozvovej premennej sme dostali ako hornú hranicu jednostranného predikčného intervalu pre priemer K budúcich meraní:

$$y_c = \hat{a} + t_{1-\alpha}(v)\hat{\sigma}\sqrt{\frac{1}{K} + \frac{1}{I \cdot J} + \frac{\bar{x}^2}{s_{xx}}}$$

Zodpovedajúcu kritickú hodnotu redukovanej stavovej premennej sme odvodili z odhadu lineárnej kalibračnej funkcie

$$x_c = \frac{y_c - \hat{a}}{\hat{b}} = t_{1-\alpha}(v)\frac{\hat{\sigma}}{\hat{b}}\sqrt{\frac{1}{K} + \frac{1}{I \cdot J} + \frac{\bar{x}^2}{s_{xx}}}$$

Nakoniec sme z kalibračnej funkcie odvodili minimálnu detegovateľnú hodnotu x_d s pravdepodobnosťou $1-\beta$, pre ktorú platí

$$x_d = \delta\frac{\hat{\sigma}}{\hat{b}}\sqrt{\frac{1}{K} + \frac{1}{I \cdot J} + \frac{\bar{x}^2}{s_{xx}}}$$

kde $\delta = \delta(v, \alpha, \beta)$ je hodnota parametra necentrality náhodnej premennej $T(v,\delta)$ s necetrálnym *t*-rozdelením so stupňami voľnosti $v = I \cdot J - 2$ a parametrom necentrality δ. Táto náhodná premenná spĺňa rovnicu

$$P[T(v,\delta) \le t_{1-\alpha}(v)] = \beta$$

3 Výpočet hodnôt parametra necentrality pre rôzne hodnoty v, α a β

Vypočítali sme hodnoty parametra necentrality pre rôzne stupne voľnosti v a pre rôzne hodnoty α a β. Výsledky sú prehľadne uvedené v nesledovných tabuľkách.

4 Záver

Výsledné tabuľky parametrov necentrality sú konštruované iba pre prípad uvedený v tomto článku a v [1]. Meracie metódy sú súčasťou riadenia kvality (pozri napríklad [2], [3], [4] a [5]).

Literatúra

[1] JANIGA, I., GARAJ, I., CISKO, P. Estimation of Minimum Detectable Value by Using Linear Calibration with Constant Standard Deviation. In *FORUM METRICUM SLOVACUM*. ISBN 80-88946-23-9. 2004, Tom VIII., pp. 68-73.

[2] GARAJ, I., JANIGA, I. *Jednostranné tolerančné medze normálneho rozdelenia s neznámou strednou hodnotou a rozptylom. One sided tolerance limits of normal distribution with unknown mean and variability.* Bratislava: STU, 2005. 214 s. ISBN 80-227-2218-9.

[3] TEREK, M., HRNČIAROVÁ, Ľ. *Štatistické riadenie kvality*. Vydavateľstvo IURA EDITION, 2004, 234 s. ISBN 80-89047-97-1.

[4] TEREK, M., HRNČIAROVÁ, Ľ. *Analýza spôsobilosti procesu*. Vydavateľstvo EKONÓM, Ekonomická univerzita v Bratislave, 2001. 205 s. ISBN 80-225-1443-8.

[5] PALENČÁR, R., RUIZ, J.M., JANIGA, I., HORNÍKOVÁ, A. *Štatistické metódy v skúšobných a kalibračných laboratóriách*. Bratislava: Grafické štúdio Ing. Peter Juriga, 2001. 380 s. ISBN 80-968449-3-8.

Tento článok vznikol s podporou grantových projektov VEGA č. 1/1247/04 Progresívne štatistické techniky a rozhodovanie v procese zlepšovania kvality a č. 1/3182/06 Zlepšovanie kvality produkcie strojárskych výrobkov pomocou moderných štatistických metód.

Kontaktné adresy autorov:

Doc. RNDr. Ivan Janiga, PhD., Katedra matematiky, SjF STU, Nám. slobody 17,
812 31 Bratislava, tel.: +421-2-57296-160, e-mail: ivan.janiga@stuba.sk

RNDr. Ivan Garaj, PhD., Ústav informatizácie, automatizácie a matematiky, FCHPT STU,
Radlinského 9, 812 37 Bratislava, tel.: +421-2-59325 297, e-mail: ivan.garaj@stuba.sk

Tabuľka 1 – Hodnoty parametra necentrality $\delta(\nu;\alpha;\beta)$ pre stupne voľnosti ν
$\alpha = 0{,}05$ **a** $\beta = 0{,}001;\ 0{,}005;\ 0{,}01;\ 0{,}025;\ 0{,}05;\ 0{,}075;\ 0{,}10;\ 0{,}15$ *a* $0{,}20$

ν \ β	0,001	0,005	0,01	0,025	0,05	0,075	0,10	0,15	0,20
1	21,0346	17,9438	16,4659	14,3281	12,5290	11,3815	10,5147	9,2021	8,1923
2	8,4619	7,3934	6,8823	6,1418	5,5159	5,1144	4,8092	4,3429	3,9790
3	6,6234	5,8416	5,4663	4,9204	4,4564	4,1569	3,9282	3,5764	3,2994
4	5,9643	5,2827	4,9546	4,4759	4,0673	3,8026	3,5999	3,2870	3,0397
5	5,6372	5,0041	4,6986	4,2521	3,8699	3,6220	3,4317	3,1375	2,9046
6	5,4447	4,8393	4,5467	4,1185	3,7516	3,5132	3,3301	3,0468	2,8221
7	5,3188	4,7310	4,4467	4,0302	3,6730	3,4408	3,2623	2,9859	2,7666
8	5,2303	4,6546	4,3760	3,9676	3,6171	3,3891	3,2139	2,9423	2,7268
9	5,1649	4,5979	4,3234	3,9210	3,5754	3,3505	3,1776	2,9096	2,6969
10	5,1145	4,5542	4,2829	3,8849	3,5430	3,3205	3,1494	2,8842	2,6736
11	5,0747	4,5196	4,2507	3,8562	3,5173	3,2966	3,1269	2,8638	2,6549
12	5,0424	4,4914	4,2245	3,8328	3,4962	3,2771	3,1085	2,8472	2,6396
13	5,0156	4,4681	4,2027	3,8134	3,4787	3,2608	3,0932	2,8333	2,6269
14	4,9932	4,4484	4,1844	3,7970	3,4640	3,2471	3,0803	2,8216	2,6161
15	4,9740	4,4317	4,1688	3,7830	3,4513	3,2353	3,0692	2,8115	2,6068
16	4,9575	4,4172	4,1553	3,7709	3,4404	3,2252	3,0596	2,8028	2,5988
17	4,9431	4,4046	4,1435	3,7603	3,4309	3,2163	3,0512	2,7952	2,5918
18	4,9304	4,3935	4,1332	3,7510	3,4225	3,2085	3,0438	2,7885	2,5856
19	4,9192	4,3836	4,1240	3,7427	3,4150	3,2015	3,0373	2,7825	2,5801
20	4,9093	4,3748	4,1158	3,7354	3,4083	3,1953	3,0314	2,7771	2,5751
21	4,9003	4,3670	4,1084	3,7287	3,4023	3,1897	3,0261	2,7723	2,5707
22	4,8922	4,3598	4,1017	3,7228	3,3969	3,1846	3,0213	2,7680	2,5667
23	4,8849	4,3534	4,0957	3,7173	3,3920	3,1800	3,0170	2,7640	2,5630
24	4,8782	4,3475	4,0902	3,7124	3,3875	3,1759	3,0130	2,7604	2,5597
25	4,8721	4,3421	4,0851	3,7078	3,3834	3,1720	3,0094	2,7571	2,5566
26	4,8665	4,3372	4,0805	3,7037	3,3796	3,1685	3,0061	2,7541	2,5538
27	4,8613	4,3326	4,0762	3,6998	3,3761	3,1652	3,0030	2,7513	2,5512
28	4,8565	4,3284	4,0723	3,6962	3,3729	3,1622	3,0001	2,7487	2,5488
29	4,8521	4,3245	4,0686	3,6929	3,3699	3,1594	2,9975	2,7463	2,5466
30	4,8480	4,3208	4,0652	3,6899	3,3671	3,1568	2,9950	2,7440	2,5445
31	4,8441	4,3174	4,0620	3,6870	3,3645	3,1544	2,9927	2,7419	2,5426
32	4,8406	4,3143	4,0591	3,6843	3,3621	3,1521	2,9906	2,7400	2,5408
33	4,8372	4,3113	4,0563	3,6818	3,3598	3,1500	2,9886	2,7381	2,5391
34	4,8341	4,3085	4,0537	3,6795	3,3577	3,1480	2,9867	2,7364	2,5375
35	4,8311	4,3059	4,0512	3,6772	3,3556	3,1461	2,9849	2,7348	2,5360
36	4,8283	4,3034	4,0489	3,6752	3,3537	3,1443	2,9832	2,7332	2,5346
37	4,8257	4,3011	4,0467	3,6732	3,3520	3,1427	2,9816	2,7318	2,5332
38	4,8232	4,2989	4,0447	3,6713	3,3503	3,1411	2,9801	2,7304	2,5320
39	4,8208	4,2968	4,0427	3,6696	3,3487	3,1396	2,9787	2,7291	2,5308
40	4,8186	4,2948	4,0409	3,6679	3,3471	3,1382	2,9774	2,7279	2,5296
41	4,8165	4,2930	4,0391	3,6663	3,3457	3,1368	2,9761	2,7267	2,5286
42	4,8144	4,2912	4,0374	3,6648	3,3443	3,1355	2,9749	2,7256	2,5275
43	4,8125	4,2895	4,0358	3,6633	3,3430	3,1343	2,9737	2,7246	2,5266
44	4,8107	4,2879	4,0343	3,6620	3,3418	3,1331	2,9726	2,7236	2,5256
45	4,8089	4,2863	4,0329	3,6607	3,3406	3,1320	2,9716	2,7226	2,5247
46	4,8073	4,2848	4,0315	3,6594	3,3394	3,1310	2,9705	2,7217	2,5239
47	4,8057	4,2834	4,0302	3,6582	3,3383	3,1299	2,9696	2,7208	2,5231
48	4,8042	4,2821	4,0289	3,6571	3,3373	3,1290	2,9687	2,7199	2,5223
49	4,8027	4,2808	4,0277	3,6560	3,3363	3,1280	2,9678	2,7191	2,5215
50	4,8013	4,2796	4,0265	3,6549	3,3354	3,1271	2,9669	2,7184	2,5208
51	4,8000	4,2784	4,0254	3,6539	3,3344	3,1263	2,9661	2,7176	2,5201
52	4,7987	4,2772	4,0243	3,6530	3,3336	3,1255	2,9653	2,7169	2,5195
53	4,7974	4,2761	4,0233	3,6520	3,3327	3,1247	2,9646	2,7162	2,5188
54	4,7963	4,2751	4,0223	3,6511	3,3319	3,1239	2,9639	2,7156	2,5182
55	4,7951	4,2740	4,0214	3,6503	3,3311	3,1232	2,9632	2,7149	2,5176
56	4,7940	4,2731	4,0204	3,6494	3,3304	3,1225	2,9625	2,7143	2,5171
57	4,7929	4,2721	4,0195	3,6486	3,3296	3,1218	2,9619	2,7137	2,5165
58	4,7919	4,2712	4,0187	3,6479	3,3289	3,1211	2,9612	2,7132	2,5160
59	4,7909	4,2703	4,0179	3,6471	3,3282	3,1205	2,9606	2,7126	2,5155
60	4,7900	4,2695	4,0171	3,6464	3,3276	3,1199	2,9600	2,7121	2,5150

Pokračovanie tabuľky 1

ν \ β	0,001	0,005	0,01	0,025	0,05	0,075	0,10	0,15	0,20
61	4,7890	4,2687	4,0163	3,6457	3,3270	3,1193	2,9595	2,7116	2,5145
62	4,7882	4,2679	4,0156	3,6450	3,3263	3,1187	2,9589	2,7111	2,5141
63	4,7873	4,2671	4,0148	3,6444	3,3257	3,1181	2,9584	2,7106	2,5136
64	4,7865	4,2664	4,0141	3,6437	3,3252	3,1176	2,9579	2,7101	2,5132
65	4,7856	4,2656	4,0135	3,6431	3,3246	3,1171	2,9574	2,7097	2,5128
66	4,7849	4,2649	4,0128	3,6425	3,3241	3,1166	2,9569	2,7092	2,5124
67	4,7841	4,2643	4,0122	3,6420	3,3236	3,1161	2,9565	2,7088	2,5120
68	4,7834	4,2636	4,0116	3,6414	3,3230	3,1156	2,9560	2,7084	2,5116
69	4,7826	4,2630	4,0110	3,6409	3,3226	3,1152	2,9556	2,7080	2,5112
70	4,7819	4,2624	4,0104	3,6403	3,3221	3,1147	2,9552	2,7076	2,5109
71	4,7813	4,2618	4,0098	3,6398	3,3216	3,1143	2,9547	2,7072	2,5105
72	4,7806	4,2612	4,0093	3,6393	3,3212	3,1139	2,9543	2,7069	2,5102
73	4,7800	4,2606	4,0087	3,6389	3,3207	3,1134	2,9540	2,7065	2,5098
74	4,7794	4,2601	4,0082	3,6384	3,3203	3,1130	2,9536	2,7062	2,5095
75	4,7788	4,2595	4,0077	3,6379	3,3199	3,1127	2,9532	2,7058	2,5092
76	4,7782	4,2590	4,0072	3,6375	3,3195	3,1123	2,9529	2,7055	2,5089
77	4,7776	4,2585	4,0067	3,6371	3,3191	3,1119	2,9525	2,7052	2,5086
78	4,7770	4,2580	4,0063	3,6366	3,3187	3,1116	2,9522	2,7049	2,5083
79	4,7765	4,2575	4,0058	3,6362	3,3183	3,1112	2,9518	2,7046	2,5080
80	4,7760	4,2571	4,0054	3,6358	3,3180	3,1109	2,9515	2,7043	2,5078
81	4,7755	4,2566	4,0050	3,6354	3,3176	3,1105	2,9512	2,7040	2,5075
82	4,7749	4,2562	4,0045	3,6351	3,3173	3,1102	2,9509	2,7037	2,5072
83	4,7745	4,2557	4,0041	3,6347	3,3169	3,1099	2,9506	2,7034	2,5070
84	4,7740	4,2553	4,0037	3,6343	3,3166	3,1096	2,9503	2,7032	2,5067
85	4,7735	4,2549	4,0033	3,6340	3,3163	3,1093	2,9500	2,7029	2,5065
86	4,7731	4,2545	4,0030	3,6336	3,3160	3,1090	2,9497	2,7026	2,5063
87	4,7726	4,2541	4,0026	3,6333	3,3157	3,1087	2,9495	2,7024	2,5060
88	4,7722	4,2537	4,0022	3,6330	3,3154	3,1084	2,9492	2,7022	2,5058
89	4,7718	4,2533	4,0019	3,6326	3,3151	3,1081	2,9489	2,7019	2,5056
90	4,7713	4,2529	4,0015	3,6323	3,3148	3,1079	2,9487	2,7017	2,5054
91	4,7709	4,2526	4,0012	3,6320	3,3145	3,1076	2,9484	2,7015	2,5052
92	4,7705	4,2522	4,0009	3,6317	3,3142	3,1074	2,9482	2,7012	2,5050
93	4,7702	4,2519	4,0005	3,6314	3,3140	3,1071	2,9480	2,7010	2,5048
94	4,7377	4,2230	3,9734	3,6068	3,2915	3,0861	2,9280	2,6827	2,4878
95	4,7698	4,2515	4,0002	3,6311	3,3137	3,1069	2,9477	2,7008	2,5046
96	4,7694	4,2512	3,9999	3,6309	3,3134	3,1066	2,9475	2,7006	2,5044
97	4,7690	4,2509	3,9996	3,6306	3,3132	3,1064	2,9473	2,7004	2,5042
98	4,7687	4,2506	3,9993	3,6303	3,3129	3,1062	2,9471	2,7002	2,5040
99	4,7683	4,2503	3,9990	3,6301	3,3127	3,1059	2,9468	2,7000	2,5038
100	4,7680	4,2500	3,9987	3,6298	3,3125	3,1057	2,9466	2,6998	2,5036
110	4,7677	4,2497	3,9985	3,6295	3,3122	3,1055	2,9464	2,6996	2,5035
120	4,7646	4,2470	3,9959	3,6273	3,3102	3,1036	2,9446	2,6979	2,5019
130	4,7621	4,2448	3,9939	3,6254	3,3084	3,1020	2,9431	2,6965	2,5006
140	4,7600	4,2429	3,9921	3,6238	3,3070	3,1006	2,9418	2,6954	2,4995
150	4,7582	4,2413	3,9906	3,6224	3,3057	3,0994	2,9407	2,6943	2,4986
200	4,7567	4,2399	3,9893	3,6212	3,3047	3,0984	2,9397	2,6935	2,4978
300	4,7512	4,2351	3,9847	3,6171	3,3009	3,0949	2,9364	2,6904	2,4949
400	4,7458	4,2303	3,9802	3,6130	3,2972	3,0914	2,9330	2,6874	2,4921
500	4,7431	4,2279	3,9779	3,6109	3,2953	3,0896	2,9314	2,6858	2,4907
600	4,7415	4,2264	3,9766	3,6097	3,2942	3,0886	2,9304	2,6849	2,4898
700	4,7404	4,2255	3,9757	3,6089	3,2934	3,0879	2,9297	2,6843	2,4893
800	4,7397	4,2248	3,9750	3,6083	3,2929	3,0874	2,9292	2,6839	2,4889
900	4,7391	4,2243	3,9746	3,6079	3,2925	3,0870	2,9289	2,6836	2,4886
1000	4,7387	4,2239	3,9742	3,6075	3,2922	3,0867	2,9286	2,6833	2,4883
1250	4,7383	4,2235	3,9739	3,6073	3,2919	3,0865	2,9284	2,6831	2,4882
1500	4,7372	4,2226	3,9730	3,6064	3,2912	3,0858	2,9277	2,6825	2,4876
2000	4,7367	4,2221	3,9725	3,6060	3,2908	3,0854	2,9274	2,6822	2,4873
2500	4,7364	4,2218	3,9723	3,6058	3,2906	3,0852	2,9272	2,6820	2,4871
3000	4,7362	4,2216	3,9721	3,6056	3,2904	3,0851	2,9271	2,6819	2,4870
∞	4,7351	4,2207	3,9712	3,6048	3,2897	3,0844	2,9264	2,6813	2,4865

Tabuľka 2 – Hodnoty parametra necentrality $\delta(\nu; \alpha; \beta)$ pre stupne voľnosti
ν $\alpha = 0{,}10$ **a** $\beta = 0{,}001;$ 0,005; 0,01; 0,025; 0,05; 0,075; 0,10; 0,15 *a* 0,20

ν \ β	0,001	0,005	0,01	0,025	0,05	0,075	0,10	0,15	0,20
1	10,6484	9,0838	8,3356	7,2533	6,3426	5,7617	5,3229	4,6584	4,1472
2	6,0940	5,3099	4,9340	4,3879	3,9247	3,6264	3,3989	3,0497	2,7755
3	5,2965	4,6458	4,3325	3,8751	3,4846	3,2316	3,0378	2,7385	2,5019
4	4,9929	4,3909	4,1002	3,6749	3,3106	3,0740	2,8924	2,6114	2,3886
5	4,8369	4,2589	3,9794	3,5698	3,2186	2,9902	2,8147	2,5428	2,3272
6	4,7427	4,1787	3,9057	3,5055	3,1619	2,9383	2,7665	2,5001	2,2887
7	4,6799	4,1250	3,8563	3,4621	3,1235	2,9031	2,7337	2,4710	2,2624
8	4,6352	4,0866	3,8208	3,4309	3,0959	2,8777	2,7100	2,4499	2,2433
9	4,6017	4,0578	3,7942	3,4074	3,0750	2,8585	2,6921	2,4339	2,2288
10	4,5757	4,0353	3,7735	3,3891	3,0587	2,8435	2,6780	2,4213	2,2174
11	4,5550	4,0174	3,7569	3,3744	3,0456	2,8314	2,6667	2,4112	2,2082
12	4,5381	4,0028	3,7433	3,3623	3,0349	2,8215	2,6575	2,4029	2,2007
13	4,5241	3,9906	3,7320	3,3523	3,0259	2,8133	2,6497	2,3960	2,1944
14	4,5122	3,9803	3,7224	3,3438	3,0183	2,8062	2,6431	2,3901	2,1890
15	4,5020	3,9715	3,7142	3,3365	3,0118	2,8002	2,6375	2,3850	2,1844
16	4,4932	3,9638	3,7071	3,3302	3,0061	2,7950	2,6326	2,3806	2,1804
17	4,4855	3,9571	3,7009	3,3247	3,0012	2,7904	2,6283	2,3767	2,1768
18	4,4788	3,9512	3,6954	3,3198	2,9968	2,7864	2,6245	2,3733	2,1737
19	4,4728	3,9460	3,6905	3,3154	2,9929	2,7828	2,6211	2,3703	2,1710
20	4,4674	3,9413	3,6862	3,3115	2,9894	2,7795	2,6181	2,3676	2,1685
21	4,4626	3,9371	3,6822	3,3081	2,9863	2,7766	2,6153	2,3651	2,1662
22	4,4582	3,9332	3,6787	3,3049	2,9834	2,7740	2,6129	2,3629	2,1642
23	4,4542	3,9298	3,6755	3,3020	2,9809	2,7716	2,6106	2,3609	2,1624
24	4,4506	3,9266	3,6725	3,2994	2,9785	2,7694	2,6086	2,3590	2,1607
25	4,4473	3,9237	3,6698	3,2970	2,9763	2,7674	2,6067	2,3573	2,1591
26	4,4443	3,9211	3,6673	3,2948	2,9743	2,7656	2,6050	2,3558	2,1577
27	4,4414	3,9186	3,6650	3,2927	2,9725	2,7639	2,6034	2,3543	2,1564
28	4,4388	3,9163	3,6629	3,2908	2,9708	2,7623	2,6019	2,3530	2,1552
29	4,4364	3,9142	3,6610	3,2891	2,9692	2,7608	2,6005	2,3517	2,1540
30	4,4342	3,9122	3,6591	3,2874	2,9677	2,7595	2,5992	2,3506	2,1530
31	4,4321	3,9104	3,6574	3,2859	2,9664	2,7582	2,5980	2,3495	2,1520
32	4,4301	3,9087	3,6558	3,2845	2,9651	2,7570	2,5969	2,3485	2,1511
33	4,4283	3,9071	3,6543	3,2831	2,9639	2,7559	2,5958	2,3476	2,1502
34	4,4266	3,9056	3,6529	3,2819	2,9627	2,7548	2,5949	2,3467	2,1494
35	4,4250	3,9042	3,6516	3,2807	2,9617	2,7538	2,5939	2,3458	2,1486
36	4,4234	3,9028	3,6503	3,2795	2,9607	2,7529	2,5931	2,3450	2,1479
37	4,4220	3,9016	3,6491	3,2785	2,9597	2,7520	2,5922	2,3443	2,1472
38	4,4206	3,9004	3,6480	3,2775	2,9588	2,7512	2,5914	2,3436	2,1466
39	4,4193	3,8992	3,6470	3,2765	2,9580	2,7504	2,5907	2,3429	2,1460
40	4,4181	3,8981	3,6460	3,2756	2,9572	2,7497	2,5900	2,3423	2,1454
41	4,4169	3,8971	3,6450	3,2748	2,9564	2,7489	2,5893	2,3417	2,1448
42	4,4158	3,8962	3,6441	3,2740	2,9557	2,7483	2,5887	2,3411	2,1443
43	4,4148	3,8952	3,6432	3,2732	2,9550	2,7476	2,5881	2,3405	2,1438
44	4,4138	3,8944	3,6424	3,2725	2,9543	2,7470	2,5875	2,3400	2,1433
45	4,4128	3,8935	3,6416	3,2718	2,9537	2,7464	2,5869	2,3395	2,1429
46	4,4119	3,8927	3,6409	3,2711	2,9531	2,7459	2,5864	2,3390	2,1424
47	4,4110	3,8919	3,6402	3,2705	2,9525	2,7453	2,5859	2,3386	2,1420
48	4,4102	3,8912	3,6395	3,2698	2,9519	2,7448	2,5854	2,3382	2,1416
49	4,4094	3,8905	3,6388	3,2692	2,9514	2,7443	2,5850	2,3377	2,1412
50	4,4086	3,8898	3,6382	3,2687	2,9509	2,7438	2,5845	2,3373	2,1409
51	4,4079	3,8891	3,6376	3,2681	2,9504	2,7434	2,5841	2,3370	2,1405
52	4,4071	3,8885	3,6370	3,2676	2,9499	2,7429	2,5837	2,3366	2,1402
53	4,4065	3,8879	3,6364	3,2671	2,9495	2,7425	2,5833	2,3362	2,1399
54	4,4058	3,8873	3,6359	3,2666	2,9490	2,7421	2,5829	2,3359	2,1395
55	4,4052	3,8868	3,6354	3,2662	2,9486	2,7417	2,5825	2,3356	2,1392
56	4,4046	3,8862	3,6349	3,2657	2,9482	2,7414	2,5822	2,3352	2,1390
57	4,4040	3,8857	3,6344	3,2653	2,9478	2,7410	2,5818	2,3349	2,1387
58	4,4034	3,8852	3,6339	3,2649	2,9475	2,7406	2,5815	2,3346	2,1384
59	4,4029	3,8848	3,6335	3,2645	2,9471	2,7403	2,5812	2,3343	2,1381
60	4,4023	3,8843	3,6330	3,2641	2,9467	2,7400	2,5809	2,3341	2,1379

Pokračovanie tabuľky 2

ν \ β	0,001	0,005	0,01	0,025	0,05	0,075	0,10	0,15	0,20
61	4,4018	3,8838	3,6326	3,2637	2,9464	2,7397	2,5806	2,3338	2,1376
62	4,4013	3,8834	3,6322	3,2633	2,9461	2,7394	2,5803	2,3335	2,1374
63	4,4009	3,8830	3,6318	3,2630	2,9458	2,7391	2,5800	2,3333	2,1372
64	4,4004	3,8826	3,6314	3,2626	2,9455	2,7388	2,5798	2,3330	2,1370
65	4,3999	3,8822	3,6311	3,2623	2,9452	2,7385	2,5795	2,3328	2,1367
66	4,3995	3,8818	3,6307	3,2620	2,9449	2,7382	2,5793	2,3326	2,1365
67	4,3991	3,8814	3,6304	3,2617	2,9446	2,7380	2,5790	2,3324	2,1363
68	4,3987	3,8811	3,6300	3,2614	2,9443	2,7377	2,5788	2,3322	2,1361
69	4,3983	3,8807	3,6297	3,2611	2,9441	2,7375	2,5785	2,3319	2,1359
70	4,3979	3,8804	3,6294	3,2608	2,9438	2,7373	2,5783	2,3317	2,1358
71	4,3975	3,8801	3,6291	3,2605	2,9436	2,7370	2,5781	2,3315	2,1356
72	4,3972	3,8797	3,6288	3,2603	2,9433	2,7368	2,5779	2,3314	2,1354
73	4,3968	3,8794	3,6285	3,2600	2,9431	2,7366	2,5777	2,3312	2,1352
74	4,3965	3,8791	3,6282	3,2598	2,9429	2,7364	2,5775	2,3310	2,1351
75	4,3961	3,8788	3,6279	3,2595	2,9426	2,7362	2,5773	2,3308	2,1349
76	4,3958	3,8785	3,6277	3,2593	2,9424	2,7360	2,5771	2,3306	2,1348
77	4,3955	3,8783	3,6274	3,2590	2,9422	2,7358	2,5769	2,3305	2,1346
78	4,3952	3,8780	3,6272	3,2588	2,9420	2,7356	2,5767	2,3303	2,1345
79	4,3949	3,8777	3,6269	3,2586	2,9418	2,7354	2,5766	2,3302	2,1343
80	4,3946	3,8775	3,6267	3,2584	2,9416	2,7352	2,5764	2,3300	2,1342
81	4,3943	3,8772	3,6264	3,2582	2,9414	2,7350	2,5762	2,3299	2,1340
82	4,3940	3,8770	3,6262	3,2580	2,9412	2,7349	2,5761	2,3297	2,1339
83	4,3937	3,8767	3,6260	3,2578	2,9410	2,7347	2,5759	2,3296	2,1338
84	4,3935	3,8765	3,6258	3,2576	2,9409	2,7345	2,5758	2,3294	2,1336
85	4,3932	3,8763	3,6256	3,2574	2,9407	2,7344	2,5756	2,3293	2,1335
86	4,3930	3,8760	3,6253	3,2572	2,9405	2,7342	2,5755	2,3292	2,1334
87	4,3927	3,8758	3,6251	3,2570	2,9404	2,7341	2,5753	2,3290	2,1333
88	4,3925	3,8756	3,6249	3,2568	2,9402	2,7339	2,5752	2,3289	2,1332
89	4,3922	3,8754	3,6247	3,2566	2,9401	2,7338	2,5750	2,3288	2,1331
90	4,3920	3,8752	3,6246	3,2565	2,9399	2,7336	2,5749	2,3287	2,1329
91	4,3918	3,8750	3,6244	3,2563	2,9397	2,7335	2,5748	2,3285	2,1328
92	4,3916	3,8748	3,6242	3,2561	2,9396	2,7333	2,5747	2,3284	2,1327
93	4,3913	3,8746	3,6240	3,2560	2,9395	2,7332	2,5745	2,3283	2,1326
94	4,3911	3,8744	3,6238	3,2558	2,9393	2,7331	2,5744	2,3282	2,1325
95	4,3909	3,8743	3,6237	3,2557	2,9392	2,7330	2,5743	2,3281	2,1324
96	4,3907	3,8741	3,6235	3,2555	2,9390	2,7328	2,5742	2,3280	2,1323
97	4,3905	3,8739	3,6233	3,2554	2,9389	2,7327	2,5740	2,3279	2,1322
98	4,3903	3,8737	3,6232	3,2552	2,9388	2,7326	2,5739	2,3278	2,1321
99	4,3901	3,8736	3,6230	3,2551	2,9387	2,7325	2,5738	2,3277	2,1320
100	4,3900	3,8734	3,6229	3,2550	2,9385	2,7324	2,5737	2,3276	2,1320
110	4,3883	3,8719	3,6215	3,2537	2,9374	2,7313	2,5727	2,3267	2,1312
120	4,3869	3,8707	3,6204	3,2527	2,9365	2,7305	2,5719	2,3260	2,1305
130	4,3857	3,8697	3,6194	3,2518	2,9357	2,7297	2,5713	2,3254	2,1299
140	4,3847	3,8688	3,6186	3,2511	2,9350	2,7291	2,5707	2,3248	2,1294
150	4,3838	3,8680	3,6178	3,2505	2,9345	2,7286	2,5702	2,3244	2,1290
200	4,3808	3,8653	3,6153	3,2482	2,9324	2,7267	2,5684	2,3228	2,1275
300	4,3778	3,8627	3,6129	3,2460	2,9304	2,7248	2,5666	2,3212	2,1261
400	4,3763	3,8614	3,6116	3,2449	2,9294	2,7239	2,5657	2,3204	2,1254
500	4,3754	3,8606	3,6109	3,2442	2,9288	2,7233	2,5652	2,3199	2,1249
600	4,3748	3,8600	3,6104	3,2437	2,9284	2,7229	2,5649	2,3196	2,1246
700	4,3744	3,8596	3,6100	3,2434	2,9281	2,7227	2,5646	2,3193	2,1244
800	4,3740	3,8594	3,6098	3,2432	2,9279	2,7225	2,5644	2,3192	2,1243
900	4,3738	3,8591	3,6095	3,2430	2,9277	2,7223	2,5643	2,3190	2,1241
1000	4,3736	3,8590	3,6094	3,2429	2,9276	2,7222	2,5642	2,3189	2,1240
1250	4,3732	3,8587	3,6091	3,2426	2,9274	2,7220	2,5639	2,3187	2,1239
1500	4,3730	3,8584	3,6089	3,2424	2,9272	2,7218	2,5638	2,3186	2,1238
2000	4,3727	3,8582	3,6086	3,2422	2,9270	2,7216	2,5636	2,3185	2,1236
2500	4,3725	3,8580	3,6085	3,2421	2,9269	2,7215	2,5635	2,3184	2,1235
3000	4,3724	3,8579	3,6084	3,2420	2,9268	2,7215	2,5635	2,3183	2,1235
∞	4,3718	3,8574	3,6079	3,2415	2,9264	2,7211	2,5631	2,3180	2,1232

Zisk zo stratifikácie

Milan Milan, Ľubica Hrnčiarová

Abstract

Sampling Theory of Surveys, that is not based on the modeling of the population, uses the final population. This paper contains a description of the quantification which measures the increase in the accuracy of the point estimations by using stratified random samplings.

1. Úvod

Medzi procedúry, ktoré umožňujú zlepšiť presnosť odhadov získaných na základe náhodných výberov v konečných základných súboroch sú dve mimoriadne dôležité:

1. Prvá predchádza realizáciu výberu. Spočíva v rozdelení základného súboru na niekoľko homogénnych skupín a rozvrhnúť výberový súbor medzi tieto skupiny, aby sa zmenšila chyba odhadu- *stratifikácia* (stratifikácia a priori).
2. Druhá sa používa v etape extrapolácie výsledkov a je založená na doplnkových štatistických informáciách – *poststratifikácia* (stratifikácia a posteriori).

2. Získanie presnosti odhadov vďaka stratifikácii

Uvažujme výberový súbor stratifikovaný podľa znaku A (kvantitatívneho alebo kvalitatívneho). Chceme odhadnúť priemer μ_k konečného základného súboru.

Bodový odhad priemeru μ_k konečného základného súboru je:

$$\overline{X}^{)} = \sum_{h=1}^{m} \frac{N_h}{N} \overline{X}_h$$

Keby sme nestratifikovali, odhadli by sme priemer μ_k konečného základného súboru pomocou výberového priemeru $\overline{X}$:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

z nestratifikovaného výberu.

Porovnáme rozptyly bodových odhadov $\overline{X}^{)}$ a $\overline{X}$, ktoré zodpovedajú výberom rovnakého rozsahu.

Zisk zo stratifikácie je

$$G = D\left(\overline{X}\right) - D\left(\overline{X}^{)}\right)$$

2.1 Rozptyl $D\left(\overline{X}\right)$ bodového odhadu priemeru konečného súboru z nestratifikovaného výberu a rozptyl $D\left(\overline{X}^{)}\right)$ bodového odhadu priemeru konečného súboru zo stratifikovaného výberu

Predpokladáme náhodný výber bez opakovania. Potom *rozptyl výberového priemeru* $D\left(\overline{X}\right)$ *z nestratifikovaného výberu je*

$$D\left(\overline{X}\right) = \frac{N - n}{N - 1} \cdot \frac{\sigma^2}{n}$$

Nevychýleným *bodovým odhadom rozptylu výberového priemeru* $D\left(\overline{X}\right)$ je

$$\hat{D}\left(\overline{X}\right) = \frac{N-n}{N-1} \cdot \frac{N-1}{N} \cdot \frac{S^2}{n} = \left(\frac{N-n}{N}\right)\frac{S^2}{n} = \left(1 - \frac{n}{N}\right)\frac{S^2}{n}$$

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

Keď súbor rozdelíme na stráta, rozptyl premennej *X* v základnom súbore možno rozdeliť na súčet dvoch členov:

- rozptyl priemerov strát (rozptyl medzi strátami) a
- priemer rozptylov strát (rozptyl vnútri strát):

$$\sigma^2 = \sum_{h=1}^{m}\frac{N_h}{N} \cdot \left(\mu_{kh} - \mu_k\right)^2 + \sum_{h=1}^{m}\frac{N_h}{N} \cdot \sigma_{kh}^2$$

Potom *rozptyl* $D\left(\overline{X}\right)$ *výberového priemeru je*

$$D\left(\overline{X}\right) = \frac{N - n}{N - 1} \cdot \frac{1}{n} \cdot \left(\sum_{h=1}^{m}\frac{N_h}{N} \cdot \left(\mu_{kh} - \mu_k\right)^2 + \sum_{h=1}^{m}\frac{N_h}{N} \cdot \sigma_{kh}^2\right).$$

Rozptyl $D\left(\overline{X}^{)}\right)$ *bodového odhadu priemeru konečného základného súboru zo stratifikovaného výberu je:*

$$D\left(\overline{X}^{)}\right) = \sum_{h=1}^{m}\frac{N_n^2}{N^2} \cdot D\left(\overline{X}_h\right) = \sum_{h=1}^{m}\frac{N_h^2}{N^2} \cdot \frac{N_h - n_h}{N_h - 1} \cdot \frac{\sigma_{kh}^2}{n_h}$$

Keď je výberový pomer v rôznych strátoch rovnaký (*stratifikovaný výberový súbor s rovnakým výberovým pomerom*):

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \ldots = \frac{n_{\mathrm{m}}}{N_{\mathrm{m}}} = \frac{n}{N} = v$$

možno vo výraze na výpočet $D\left(\overline{X}\right)$ nahradiť $\frac{n_h}{N_h}$ výrazom $\frac{n}{N}$:

$$D\left(\overline{X}\right)=\sum_{h=1}^{m}\frac{N_h^2}{N^2}\cdot D\left(\overline{X}_h\right)=\sum_{h=1}^{m}\frac{N_h}{N^2}\cdot\frac{N_h-n_h}{N_h}\cdot\frac{N_h}{N_h-1}\cdot\frac{N_h}{n_h}\cdot\sigma_{kh}^2=$$

$$=\sum_{h=1}^{m}\frac{N_n}{N^2}\left(1-\frac{n}{N}\right)\cdot\frac{N_h}{N_h-1}\cdot\frac{N}{n}\cdot\sigma_{kh}^2=\sum_{h=1}^{m}\frac{N_h}{N^2}\cdot\frac{N-n}{N}\cdot\frac{N_h}{N_h-1}\cdot\frac{N}{n}\cdot\sigma_{kh}^2=$$

$$=\frac{N-n}{N}\cdot\frac{1}{n}\cdot\sum_{h=1}^{m}\frac{N_h}{N}\cdot\frac{N_h}{N_h-1}\cdot\sigma_{kh}^2=$$

$$=\frac{1-v}{n}\cdot\sum_{h=1}^{m}\frac{N_h}{N}\cdot\frac{N_h}{N_h-1}\cdot\sigma_{kh}^2$$

Vo výraze je σ_{kh}^2 neznáma, možno ju odhadnúť pomocou:

$$\frac{N_h-1}{N_h}\cdot S_h^2,$$

kde

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

Nakoniec dostaneme *bodový odhad rozptylu* $D\left(\overline{X}\right)$ *bodového odhadu priemeru konečného základného súboru zo stratifikovaného výberu:*

$$\hat{D}\left(\overline{X}\right)=\frac{1-v}{n}\cdot\sum_{h=1}^{m}\frac{N_h}{N}\cdot\frac{N_h}{N_h-1}\cdot S_h^2$$

V prípade reprezentatívneho základného súboru (Keď N a N_h sú veľké možno položiť:

$\frac{1}{N-1}\approx\frac{1}{N}$ a $\frac{1}{N_h-1}\approx\frac{1}{N_h}$.

Potom:

$$D\left(\overline{X}\right)\approx\frac{N-n}{N-1}\cdot\frac{1}{n}\cdot\left(\sum_{h=1}^{m}\frac{N_h}{N}\cdot\left(\mu_{kh}-\mu_k\right)^2+\sum_{h=1}^{m}\frac{N_h}{N}\cdot\sigma_{kh}^2\right)=$$

$$=\frac{1}{N-1}\cdot\frac{N-n}{n}\cdot\left(\sum_{h=1}^{m}\frac{N_h}{N}\cdot\left(\mu_{kh}-\mu_k\right)^2+\sum_{h=1}^{m}\frac{N_h}{N}\cdot\sigma_{kh}^2\right)=\quad,$$

$$=\frac{1}{N}\cdot\frac{N-n}{n}\cdot\left(\sum_{h=1}^{m}\frac{N_h}{N}\cdot\left(\mu_{kh}-\mu_k\right)^2+\sum_{h=1}^{m}\frac{N_h}{N}\cdot\sigma_{kh}^2\right)=$$

$$=\frac{1-v}{n}\cdot\left(\sum_{h=1}^{m}\frac{N_h}{N}\left(\mu_{kh}-\mu_k\right)^2+\sum_{h=1}^{m}\frac{N_h}{N}\cdot\sigma_{kh}^2\right)$$

kde

$\left(\sum_{h=1}^{m}\frac{N_h}{N}\left(\mu_{kh}-\mu_k\right)^2+\sum_{h=1}^{m}\frac{N_h}{N}\cdot\sigma_{kh}^2\right)=\sigma^2$ je celková variabilita

$$D\left(\overline{X}^{)}\right) \approx \frac{1-v}{n} \cdot \sum_{h=1}^{m} \frac{N_h}{N} \cdot \sigma_{kh}^2$$

Potom zisk zo stratifikácie je

$$G = D\left(\overline{X}\right) - D\left(\overline{X}^{)}\right) \approx \sum_{h=1}^{m} \frac{N_h}{N} \cdot \left(\mu_{kh} - \mu_k\right)^2 ,$$

kde

$\sum_{h=1}^{m} \frac{N_h}{N} \cdot \left(\mu_{kh} - \mu_k\right)^2$ je rozptyl priemerov strát(rozptyl medzi strátami).

Relatívny zisk je

$$\frac{D\left(\overline{X}\right) - D\left(\overline{X}^{)}\right)}{D\left(\overline{X}\right)} \approx \frac{\sum_{h=1}^{m} \frac{N_h}{N} \cdot \left(\mu_{kh} - \mu_k\right)^2}{\sigma^2} = \eta_{X/A}^2$$

$\left(\sum_{h=1}^{m} \frac{N_h}{N} \left(\mu_{kh} - \mu_k\right)^2 + \sum_{h=1}^{m} \frac{N_h}{N} \cdot \sigma_{kh}^2\right) = \sigma^2$ je celková variabilita.

Štvorec konečného pomeru (pomer determinácie): $\eta_{X/A}^2 = \frac{vysvetlená \quad variabilita}{celková \quad variabilita}$

Všeobecne platí: $\eta_{Y/X}^2 \neq \eta_{X/Y}^2$.

Relatívny zisk sa teda všeobecne rovná pomeru determinácie $\eta_{X/A}^2$, kde *A* znak použitý ako kritérium stratifikácie.

Zisk zo stratifikácie je teda tým väčší, čím je väčšia korelácia X a A. Keď $\eta_{X/A}^2 = 0$ (rozptyl medzi strátami je nulový), zisk zo stratifikácie je nulový. Keď $\eta_{X/A}^2 = 1$, odhad je úplne presný a rozptyl vnútri každého stráta je nulový.

3. Záver

Vždy je záujem stratifikovať. Aj keď nepoznáme smerodajné odchýlky študovanej premennej v strátoch a nemožno stanoviť optimálne rozdelenie výberového súboru, aj tak sa preferuje stratifikácia s rovnakým výberovým pomerom (reprezentatívny stratifikovaný výber) pred absenciou stratifikácie.

Zisk zo stratifikácie je tým väčší, čím je väčšia korelácia študovanej premennej s kritériom stratifikácie. Inak, čím sa stráta z hľadiska študovanej premennej viac navzájom líšia.

Pre ilustráciu uvedieme výsledky jednej štúdie (Hansen – Hurwitz).Výsledky v tab. 1 sa týkajú prieskumu priemyselných podnikov stratifikovaných podľa obratu (kritérium stratifikácie). Porovnáva sa variačný koeficient priemerných miezd (priemerná mzda je študovanou premennou).

Tab. 1 Porovnanie presnosti bodového odhadu priemerných miezd

Strátum h	Rozsah stráta N_h	Žiadna statifikácia N	Stratifikácia s rovnakým výberovým pomerom n_h	Optimálny výberový súbor n_h	Rozdelenie úmerné obratu firmy n_h
1	538		15	244	286
2	4753		131	288	409
3	30964		854	468	305
Celkom	36258	1000	1000	1000	1000
Variačný koeficient bodového odhadu	$\frac{\sigma_{\bar{X}}}{\mu_k}$	9,9%	7,1%	3,0%	3,3%

Zdroj: Hansen – Hurwitz

Poďakovanie
Tento príspevok vznikol s prispením grantovej agentúry VEGA v rámci projektov číslo 1/1247/04 Progresívne štatistické techniky a rozhodovanie v procese zlepšovania kvality a číslo 1/3182/06 Zlepšovanie kvality produkcie strojárskych výrobkov pomocou štatistických metód.

Použitá literatúra

1. Grais B.: Méthodes statistiques. Techniques statistiques 2. Paris, Dunod 2000
2. Tillé Y.: Théorie des sondages. Paris, Dunod 2001
3. Cochran, W. G.: Sampling Techniques. New York, Wiley and Sons 1977

Kontaktná adresa

doc. Ing. Ľubica Hrnčiarová, PhD., Katedra štatistiky FHI EU v Bratislave, Dolnozemská cesta 1, 852 35 Bratislava, tel: 02-67295731; e-mail: hrnciaro@dec.euba.sk

doc. Ing. Milan Terek, PhD., Katedra štatistiky FHI EU v Bratislave, Dolnozemská cesta 1, 852 35 Bratislava, tel: 02-67295713; e-mail: terek@dec.euba.sk

Zhluková analýza produkčných ukazovateľov stavebných podnikov právnej formy s. r. o.

Samuel Korónyi

Abstract: The paper provides k-means clustering results of basic absolute economic production indicators of Slovak construction companies (ltd. companies) from the year 2001 (output: value added, inputs: sum of assets and average number of employee) including graphical presentation in scatter plots. Applied cluster method clearly defines three distinctive clusters.

Úvod

Cieľom príspevku je prezentovať výsledky aplikovania zhlukovacej metódy *k*-priemerov na vybrané absolútne ekonomické produkčné ukazovatele slovenských stavebných podnikov právnej formy s. r. o.

Popis metódy

Zhlukovacia metóda *k*-priemerov je vhodná pre vytvorenie malého počtu zhlukov z veľkého počtu objektov. Vyžaduje spojité dáta bez odľahlých údajov. Cieľ metódy je rozdeliť objekty v priestore premenných do zhlukov tak, aby súčet štvorcov v rámci zhlukov bol minimálny. Pretože počet možných usporiadaní je veľký, je nereálne očakávať najlepšie riešenie, skôr ide o lokálne optimum. Je to riešenie, v ktorom ľubovoľný presun objektu z jedného zhluku do druhého nezmenší súčet štvorcov v zhluku. Nevýhodou metódy je to, že výsledné riešenie závisí od počiatočnej konfigurácie. Preto je vhodné postup urobiť pre rôzne počiatočné konfigurácie a vybrať optimálne riešenie. Užívateľ zadá počet zhlukov, ktoré majú byť nájdené. Algoritmus najprv nájde centrá zhlukov. Potom rozdelí dáta do im priradených zhruba sférických zhlukov. V ďalšom kroku určí nové centra zhlukov a postup opakuje. Pri výpočte sú dáta štandardizované.

Systém NCSS náhodne zvolí príslušnosť bodu k zhluku. Táto zostava je optimalizovaná algoritmom *k*- priemerov. Použitie niekoľkých náhodných štartovacích konfigurácií podstatne zvýši pravdepodobnosť nájdenia globálneho optimálneho riešenia pre daný počet zhlukov.

Formálne metóda rozdelí N dátových bodov do K disjunktných podmnožín S_j, ktoré obsahujú N_j bodov tak, aby sa minimalizovala suma štvorcov

$$J_K = \sum_{j=1}^{K} \sum_{n \in S_j} \left| x_n - \mu_j \right|^2$$

kde $\boldsymbol{x}_n$ je vektor reprezentujúci n-tý bod a $\boldsymbol{\mu}_j$ je ťažisko podmnožiny S_j.
Percento rozptylu je definované ako

$$\frac{J_K}{J_1}.$$

Väčšinou je potrebné dvakrát urobiť analýzu, prvýkrát pre určenie optimálneho počtu zhlukov a druhýkrát pre informáciu o nájdených zhlukoch. Je možné meniť počet náhodných počiatočných konfigurácií pre vyskúšanie postupu pre hodnoty od minimálneho do maximálneho počtu zhlukov. Tiež sa nastavuje maximálny počet iterácií pred ukončením algoritmu. Algoritmus *k*-priemerov je typická zhlukovacia metóda. My dopredu nevieme koľko zhlukov obsahujú dáta, ale hľadáme zhluky podobných podnikov.

Dáta

Zhlukovej analýze boli podrobené vybrané absolútne produkčné ukazovatele anonymných 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

Ďalej sú uvedené výstupy zo systému NCSS. Z prvej časti reportu (rozptyl pri rôznom počte zhlukov) je zrejmé, že optimálny počet zhlukov je 3 alebo 4.

K-Means Cluster Analysis Report

Minimum Iteration Section

Iteration No.	No. of Clusters	Percent of Variation	Bar Chart of Percent
3	2	52.79	|||||||||||||||
4	3	34.97	||||||||||
8	4	27.77	||||||||
12	5	22.84	|||||||
15	6	20.10	|||||||

Pre získanie celkového prehľadu o nájdených zhlukoch je určená druhá časť reportu, ktorá poskytuje priemery a smerodajné odchýlky pre jednotlivé zhluky. Zhluk 3 v zásade obsahuje väčšie podniky v sledovaných ukazovateľoch, zhluk číslo 2 naopak tie najmenšie a prvý zhluk jadro údajov – stredne veľké podniky. S rastúcou veľkosťou podnikov de jure klesá ich relatívny podiel. Prejavuje sa to aj pri zhlukoch.

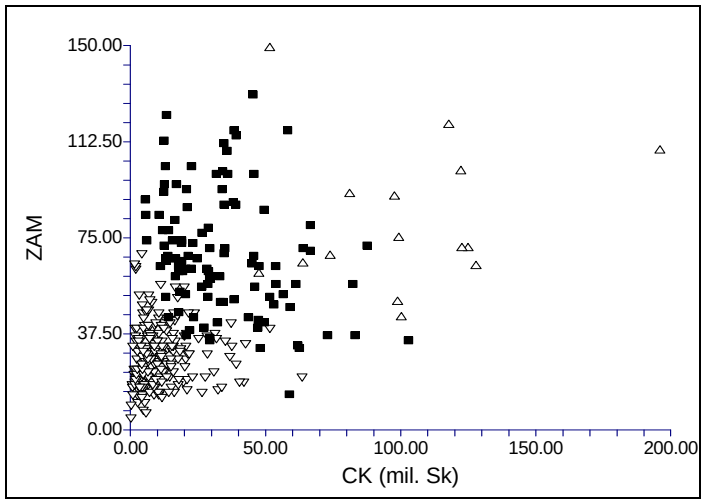
Cluster Means

Variables	Cluster1	Cluster2	Cluster3
ZAM	68	29	82
CK	33	11	102
PH	19	7	51
Count	107	239	15

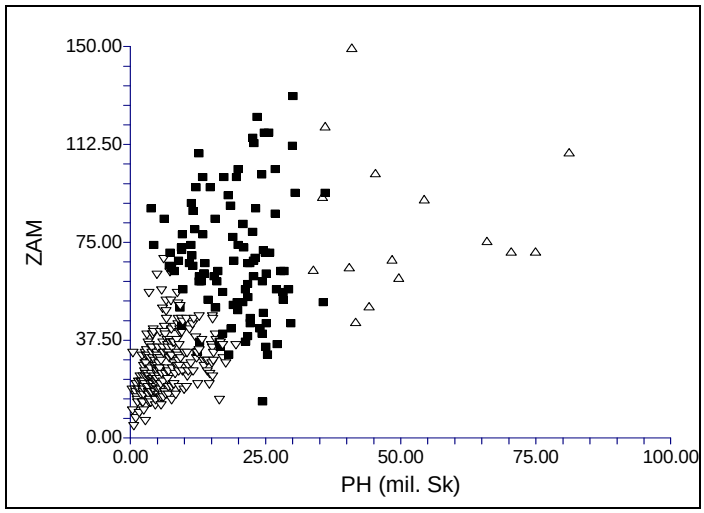
Cluster Standard Deviations

Variables	Cluster1	Cluster2	Cluster3
ZAM	23.3	11.2	28.2
CK	19.5	9.6	37.2
PH	7.0	4.0	15.2
Count	107	239	15

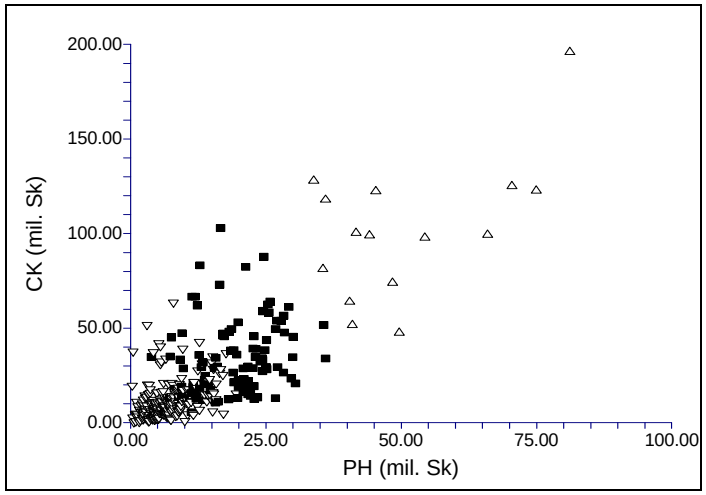
Pre grafické zobrazenie sú vhodné dvojrozmerné bodové grafy. Ich nevýhodou je možné prekrývanie zhlukov a tým následné skreslenie skutočnej situácie. Na grafoch 1 až 3 sú postupne uvedené grafy s dvojicami ukazovateľov. Prekrývanie zhlukov je len čiastočné.



Graf 1 Rozptylový graf ukazovateľov celkového kapitálu a počtu zamestnancov s nájdenými zhlukmi

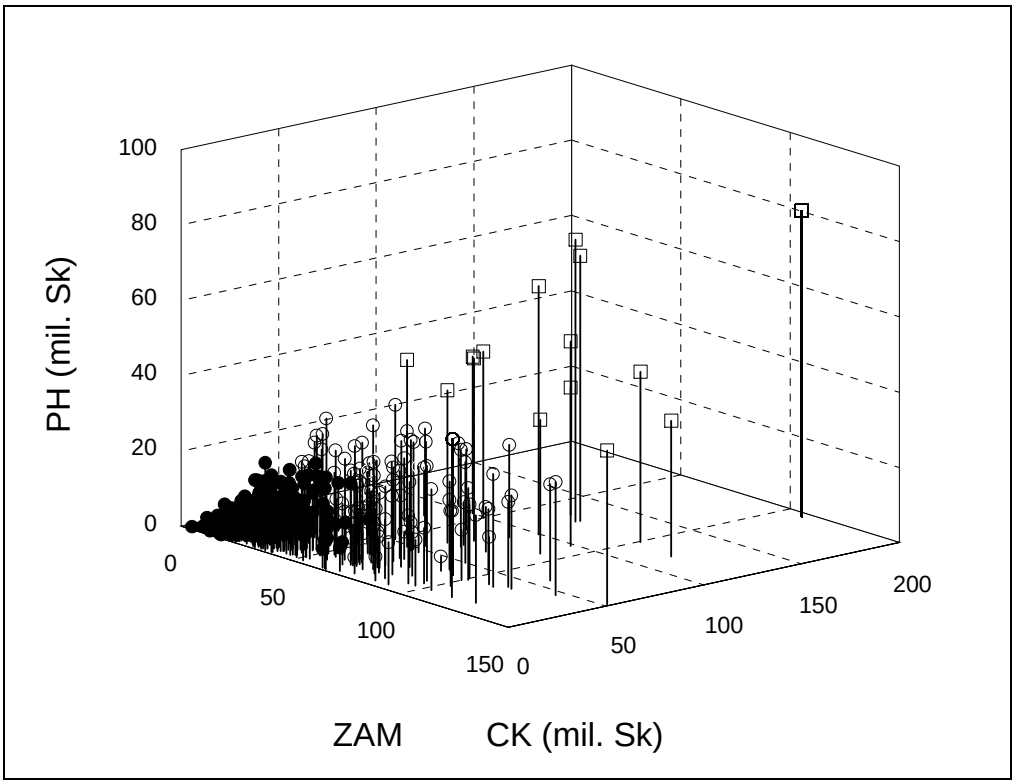


Graf 2 Rozptylový graf ukazovateľov pridanej hodnoty a počtu zamestnancov s nájdenými zhlukmi



Graf 3 Rozptylový graf ukazovateľov pridanej hodnoty a celkového kapitálu s nájdenými zhlukmi

Pre priestorový obraz sa používa trojrozmerný rozptylový graf, ktorý by mal odstrániť prípadné prekrývanie zhlukov. Jeho nevýhoda je v tom, že niektorí jednotlivci majú problémy pre vnímanie jeho hĺbky. Na grafe 4 je trojrozmerná reprezentácia zhlukov. Na grafe je zreteľný odstup a poloha všetkých troch nájdených zhlukov.



Graf 4 Rozptylový graf ukazovateľov celkového kapitálu, počtu zamestnancov a pridanej hodnoty s nájdenými zhlukmi

Záver

Štatistika poskytuje širokú paletu dostupných metód za predpokladu použitia štatistického softwaru a zhluková analýza poskytuje cenné informácie o štruktúre jedno a viacrozmerných dát. Je to jedna z mála univerzálnych štatistických metód v tom zmysle, že môže byť použitá pre výbery aj základné súbory. Podstatné je, aby nájdené zhluky boli jasne interpretovateľné, čo spĺňa náš súbor.

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

VÁHY POLYNOMICKÝCH KĹZAVÝCH PRIEMEROV

Karol Pastor

Weights in polynomial moving average formulas (Abstract). Moving average formulas are widely used for smoothing the time series, identification of their trend components or graduation of the life tables. The purpose for introduction of polynomial formulas is to avoid the distortion of a polynomial during this process. On the other hand, the weights of the 3rd-order polynomial moving average formulas minimize the error-reducing index, as well. The paper presents proofs of mentioned properties of such formulas.

1. Úvod a označenie

Kĺzavé priemery patria k štandardnej výbave analýzy časových radov. S úspechom sa využívajú najmä pri identifikácii trendovej zložky radu či pri graduácii úmrtnostných tabuliek. Popri jednoduchých kĺzavých priemeroch sa možno stretnúť s viacerými typmi vážených kĺzavých priemerov, líšiacich sa podľa vzniku: chronologické kĺzavé priemery, sumačné formuly, polynomické a optimálne kĺzavé priemery. Niekedy sa používajú aj tzv. predikčné kĺzavé priemery, ktoré vyrovnanú hodnotu priradia nie do stredu, ale na koniec skúmaného intervalu (takto ich počíta aj napr. Excel), tými sa však tu nebudeme zaoberať. Kĺzavé priemery vhodne zvolenej dĺžky dobre vystihujú trend bez toho, aby predurčovali jeho tvar, redukujú nepravidelnosti, eliminujú náhodné poruchy. Na druhej strane, nevyrovnávajú koncové hodnoty a neumožňujú predpovede (pokiaľ nepridáme osobitné formuly alebo predpoklady). Ďalšou nevýhodou je, že majú tendenciu vyrovnávať aj dostatočne hladký priebeh a tak spôsobovať skreslenie (*distortion*, viď nižšie).

Nech u_x je funkcia celočíselného argumentu x, ktorú budeme interpretovať ako skutočnú (neznámu) hodnotu radu v čase x a nech v_x je jeho pozorovaná hodnota zaťažená náhodnou chybou ε_x, t.j. $v_x = u_x + \varepsilon_x$. Označme $V_x = \hat{u}_x$ odhad u_x urobený na základe pozorovanej hodnoty v_x pomocou kĺzavého priemeru dĺžky $2r + 1$ s váhami K_i, $i = -2r, -2r+1, ..., 2r$, t.j. V_x je vyrovnaná (graduovaná) hodnota. Potom V_x možno písať ako

$$V_x = \sum_{i=-r}^{r} K_i \; v_{x+i} \tag{1}$$

O váhach K_i predpokladáme, že spĺňajú podmienky

$$\sum_{i=-r}^{r} K_i = 1, \qquad K_{-i} = K_i . \tag{2}$$

Pre jednoduché kĺzavé priemery je $K_i = 1/(2r+1)$, $i = -2r, -2r+1, ..., 2r$. Aby sa predišlo skresleniu, zavádzajú sa vhodné kombinácie jednoduchých kĺzavých priemerov alebo sa vychádza z tvaru, ktorý apriórne skreslenie polynómu vylučuje.

Takými sú aj polynomické kĺzavé priemery. Vzniknú tak, že cez $2r+1$ hodnôt preložíme polynóm vopred zvoleného stupňa. Jednoduché kĺzavé priemery sú tak vlastne polynomickými stupňa 0. V tejto práci budeme skúmať iba polynomické kĺzavé priemery stupňa 2 alebo 3 (váhy majú v oboch prípadoch rovnaký tvar). Takým priemerom dĺžky 7 je napr. i formula, ktorú používa ŠÚ SR pri výpočte úmrtnostných tabuliek (hoci metodika [8] na to neupozorňuje), konkrétne

$$V_x = [105\, v_x + 90\, (v_{x-1} + v_{x+1}) + 45\, (v_{x-2} + v_{x+2}) - 30\, (v_{x-3} + v_{x+3})] / 315 .$$

Odvodenie a tvar váh polynomických kĺzavých priemerov možno nájsť napr. v učebniciach [3], [10] alebo v monografii [7], no iba pre niekoľko konkrétnych prípadov. Odvodenie váh vo všeobecnom prípade uvádza [5], avšak v odlišnom, hoci ekvivalentnom tvare ako uvádzame tu. Problematike polynomických kĺzavých priemerov sa venujú aj diplomové práce [6] a [11].

Aby sme sa vedeli orientovať medzi množstvom rôznych formúl, zavádzajú sa rôzne kritériá kvality, menovite *index redukcie chyby (error-reducing index)* Φ_E, *index hladkosti* Φ_s , *index odrezávania vĺn* Φ_W (viď [4], [1], [9]). Index redukcie chyby porovnáva smerodajnú odchýlku vyrovnaného a nevyrovnaného radu,

$$\Phi_E = \sqrt{\frac{\mathrm{var}(\sum K_i\, \varepsilon_{x-i})}{\mathrm{var}(\varepsilon_x)}} = \sqrt{\sum_{i=-r}^{r} K_i^{\,2}} \qquad (3)$$

V učebnici [1] je uvedený všeobecný tvar váh, ktoré minimalizujú Φ_E, nie je však dokázaný. Cieľom tejto práce je doplniť túto medzeru a tiež ukázať, že jedinými kĺzavými priemermi minimalizujúcimi Φ_E sú polynomické kĺzavé priemery stupňa 2 alebo 3. Táto vlastnosť pravdepodobne je známa už dlhšie, no v dostupnej literatúre sa nedokazuje, ba ani nespomína.

2. Tvar váh polynomického kĺzavého priemeru

Veta 1 udáva tvar váh polynomických kĺzavých priemerov vo všeobecnom prípade. Pri dôkaze sa využíva nasledovná lema, ktorá sa dokazuje indukciou.

Lema 1: Pre každé prirodzené číslo r platí:

$$\sum_{i=1}^{r} i^2 = r\,(r+1)\,(2r+1) / 6 \qquad (4)$$

$$\sum_{i=1}^{r} i^4 = r\,(r+1)\,(2r+1)\,(3r^2+3r-1) / 30 \qquad (5)$$

Veta 1. Nech K_i, $i = -r, -r+1, \ldots, r$, $K_i = K_{-i}$ sú váhy polynomických kĺzavých priemerov stupňa 2 alebo 3 a dĺžky $2r+1$. Potom váhy K_i ($i = 0, 1, \ldots, r$) majú tvar

$$K_i = \frac{3\left(3r^2+3r-1\right)-15i^2}{(2r-1)(2r+1)(2r+3)} \qquad (6)$$

Dokaz. Nech v_x je pozorovaná hodnota pre pevné x. Za vyrovnanú

(graduovanú) hodnotu V_x vezmeme hodnotu ležiacu na kubickej parabole preloženej $2r+1$ bodmi ($r \geq 2$) okolo bodu v_x pre $i = 0$,

$$\hat{v}_{x+i} = \sum_{j=0}^{3} b_j i^j \quad , \quad i = -2r, -2r+1, ..., 2r \; ; \tag{7}$$

$$V_x = \hat{v}_x = b_0. \tag{8}$$

Koeficienty b_0, b_1, b_2, b_3 sa získajú metódou najmenších štvorcov, t.j. minimalizáciou výrazu

$$\sum_{i=-r}^{r} \left(v_{x+i} - \sum_{j=0}^{3} \beta_j i^j \right)^2 .$$

Príslušná sústava kánonických rovníc má tvar

$$\begin{aligned}
b_0 \sum i^0 \quad &+ b_2 \sum i^2 \quad &= \sum \; v_{x+i} \\
b_1 \sum i^2 \quad &+ b_3 \sum i^4 \; &= \sum i \; v_{x+i} \\
b_0 \sum i^2 \quad &+ b_2 \sum i^4 \quad &= \sum i^2 \; v_{x+i} \\
b_1 \sum i^4 \quad &+ b_3 \sum i^6 = \sum i^3 \; v_{x+i}
\end{aligned}$$

(lebo súčet nepárnych mocnín i je vždy nula). Odtiaľ vyjadríme b_0, pričom podľa (8)

$$V_x = b_0 = \frac{\sum i^4 \sum v_{x+i} - \sum i^2 \sum i^2 v_{x+i}}{\sum i^0 \sum i^4 - \left(\sum i^2 \right)^2} .$$

Po dosadení podľa lemy 1 (ide o dvojnásobok súm tam uvedených) dostávame

$$V_x = b_0 = \frac{\frac{2}{30} r(r+1)(2r+1)(3r^2+3r-1) \sum v_{x+i} - \frac{2}{6} r(r+1)(2r+1) \sum i^2 v_{x+i}}{(2r+1) \cdot \frac{2}{30} r(r+1)(2r+1)(3r^2+3r-1) - \left[\frac{2}{6} r(r+1)(2r+1) \right]^2} ,$$

čo po vykrátení dáva tvar

$$V_x = b_0 = \frac{3(r^2+3r-1) \sum v_{x+i} - 15 \sum i^2 v_{x+i}}{(2r-1)(2r+1)(2r+3)} . \tag{9}$$

Porovnaním koeficientov pri jednotlivých hodnotách v_{x+i} pre $i = -r, -r+1, ..., r$, pre ktoré naviac platí $K_{-i} = K_i$, čo bolo treba dokázať.

3. Skreslenie polynómu 3. stupňa

Ak $v_x = u_x$ ležia na parabole, t.j. dostatočne hladkej krivke, tak po vyrovnaní jednoduchým kĺzavým priemerom sú hodnoty $V_x \neq u_x$. Tento fakt sa volá skreslenie krivky.

Veta 2. (cf. [2]) Aby kĺzavý priemer (1) dĺžky $2r+1$ spĺňajúci podmienky (2) nespôsoboval skreslenie polynómu do 3. stupňa, musí platiť

$$\sum_{i=0}^{r} i^2 \, K_i = 0. \tag{10}$$

Dôkaz. Nech hodnoty v_x ležia na parabole 3. stupňa, t.j. pre nejaké a, b, c, d platí

$$v_{x+i} = a + b\,(x+i) + c\,(x+i)^2 + d\,(x+i)^3 =$$
$$= a + bx + bi + cx^2 + 2cxi + ci^2 + dx^3 + 3dx^2 i + 3dxi^2 + di^3 .$$

Odtiaľ

$$v_{x+i} + v_{x-i} = 2\,v_x + 2\,(c + 3dx)\,i^2. \tag{11}$$

Preto pre vyrovnanú hodnotu V_x platí

$$V_x = \sum_{i=-r}^{r} K_i \; v_{x+i} = K_0\,v_x + \sum_{i=1}^{r} K_i \; (v_{x+i} + v_{x-i}) =$$

$$= K_0\,v_x + \sum_{i=1}^{r} K_i \; [2\,v_x + 2\,(c + 3dx)\,i^2] =$$

$$= (K_0 + 2\sum_{i=1}^{r} K_i \;)\,v_x + 2\,(c + 3dx) \sum_{i=1}^{r} K_i \; i^2 . \tag{12}$$

Pretože výraz v prvej zátvorke v (12) je 1, parabola zostane bez skreslenia ($V_x = v_x$) vtedy a len vtedy, ak druhý sčítanec v (12) bude rovný nule pre každé x. Veta je dokázaná.

Polynomické kĺzavé priemery 2. a 3. stupňa nespôsobujú skreslenie polynómu 2. resp. 3. stupňa, čo vyplýva z ich konštrukcie, alebo z dosadenia do (10) alebo aj z nasledujúcej vety.

4. Minimalizácia indexu redukcie chýb

Veta 3. Nech K_i, $i = -r, -r+1, ..., r$, $K_i = K_{-i}$ sú váhy kĺzavých priemerov dĺžky $2r+1$ nespôsobujúcich skreslenie polynómu 3. stupňa, ktoré minimalizujú index redukcie chýb Φ_E . Potom váhy K_i ($i = 0, 1, ..., r$) majú tvar

$$K_i = \frac{3\left(3r^2 + 3r - 1\right) - 15\,i^2}{(2r-1)(2r+1)(2r+3)} \; . \tag{13}$$

Dôkaz. Podľa predpokladov váhy K_i spĺňajú podmienky

$$K_0 + 2\sum_{i=1}^{r} K_i \;\; = \;\; 1 \tag{14}$$

$$\sum_{i=1}^{r} i^2 K_i \;=\; 0 \qquad (15)$$

a minimalizujú index redukcie chýb Φ_E , t.j. aj jeho druhú mocninu

$$\Phi_E{}^2 = K_0{}^2 + 2\sum_{i=1}^{r} K_i{}^2 \;. \qquad (16)$$

Je to úloha na viazaný extrém s dvoma ohraničeniami. Príslušná Lagrangeova funkcia má tvar

$$L = K_0^2 + 2K_1^2 + 2K_2^2 + ... + 2K_r^2 + \lambda\left(K_0 + 2\sum_{i=1}^{r} K_i - 1\right) + \gamma\left(\sum_{i=1}^{r} i^2 K_i\right)$$

Nutné podmienky na viazané minimum sú potom tvaru:

$$\frac{\partial L}{\partial K_0} = 2K_0 + \lambda = 0 \qquad \frac{\partial L}{\partial K_i} = 4K_i + 2\lambda + \gamma i^2 = 0\,, \quad i = 1, ..., r$$

$$\frac{\partial L}{\partial \lambda} = K_0 + 2\sum_{i=1}^{r} K_i - 1 = 0 \qquad \frac{\partial L}{\partial \gamma} = \sum_{i=1}^{r} i^2 K_i = 0$$

Je to sústava $r + 3$ lineárnych rovníc o $r + 3$ neznámych. Z tretej podmienky vyjadríme K_0 a dosadíme do prvej. Dostávame $\lambda = -2(1 - 2\sum_{i=1}^{r} K_i)$. Dosadením λ do druhej dostávame

$$4\,K_i \;-\; 4\,(1 - 2\sum_{i=1}^{r} K_i) \;+\; \gamma\, i^2 \;= 0 \qquad (17)$$

a sčítaním pre $i = 1, ..., r$

$$4\,\sum_{i=1}^{r} K_i \;-\; 4\,r\,(1 - 2\sum_{i=1}^{r} K_i) + \gamma \sum_{i=1}^{r} i^2 \;= 0\,,$$

odkiaľ

$$\gamma \sum_{i=1}^{r} i^2 \;=\; 4\,r - 4\,(1 + 2r)\,\sum_{i=1}^{r} K_i \;.$$

Po vyjadrení γ a dosadení do (14) dostávame

$$4\,K_i \;-\; 4\,(1 - 2\sum_{i=1}^{r} K_i) \;+ [4\,r - 4\,(1 + 2r)\,\sum_{i=1}^{r} K_i\,]\, i^2 \Big/ \sum_{i=1}^{r} i^2 \;= 0\,. \qquad (18)$$

V ďalšom kroku vynásobíme tieto rovnice výrazom j^2 a sčítame pre všetky j

$$4\,\sum_{i=1}^{r} i^2 K_i - 4\,\sum_{i=1}^{r} i^2 \;+ 8\,\sum_{i=1}^{r} K_i \;\sum_{i=1}^{r} i^2 \;+\; [4\,r - 4\,(1 + 2r)\,\sum_{i=1}^{r} K_i\,]\,\sum_{i=1}^{r} i^4 \Big/ \sum_{i=1}^{r} i^2 \;=\; 0\,.$$

Vykrátime štyrmi a následne využijeme štvrtú z nutných podmienok, čím vypadne

prvý sčítanec. Ďalej využijeme lemu 1. Dostávame

$$-[r(r+1)(2r+1)/6] + 2[r(r+1)(2r+1)/6]\sum_{i=1}^{r} K_i \; + \; [r-(1+2r)\sum_{i=1}^{r} K_i\,][(3r+3r-1)/5] = 0$$

a po postupných úpravách

$$\{[r(r+1)(2r+1)/3] - (1+2r)[(3r^2+3r-1)/5]\}\sum_{i=1}^{r} K_i = [r(r+1)(2r+1)/6] - r[(3r^2+3r-1)/5]$$

$$\{(2r+1)[-4r^2-4r+3]/15\}\sum_{i=1}^{r} K_i = r[10r^2+15r+5-18r^2-18r+6]/30$$

$$(2r+1)(2r-1)(2r+3)\sum_{i=1}^{r} K_i = r(8r^2+3r-11)/2$$

$$\sum_{i=1}^{r} K_i = r(8r^2+3r-11)/2(2r-1)(2r+1)(2r+3) \qquad (19)$$

Dosadením (19) do tretej podmienky a úpravou máme

$$K_0 = 1 - 2\sum_{i=1}^{r} K_i = [(2r+1)(2r-1)(2r+3) - r(8r^2+3r-11)]/(2r-1)(2r+1)(2r+3)$$

$$K_0 = (9r^2+9r-3)/(2r-1)(2r+1)(2r+3) \qquad (20)$$

čo pre $i = 0$ je tvar (13). Teraz dokážeme tvrdenie pre $i = 1, ..., r$. Vzťah (18) možno podľa lemy 1 písať v tvare

$$K_i = (1-2\sum_{i=1}^{r} K_i\,) - [r-(1+2r)\sum_{i=1}^{r} K_i\,]\,i^2/[r(r+1)(2r+1)/6] \qquad (21)$$

Prvý sčítanec v (21), nezávislý od i, podľa (20) už má tvar požadovaný tvrdením vety. Hranatá zátvorka v druhom sčítanci v (21) má podľa (19) tvar

$$[2r(2r-1)(2r+1)(2r+3) - (1+2r)r(8r^2+3r-11)]/2(2r-1)(2r+1)(2r+3) =$$

$$= r(2r+1)[8r^2+8r-6-8r^2-3r+11)]/2(2r-1)(2r+1)(2r+3) =$$

$$= r(2r+1)(5r+5)/2(2r-1)(2r+1)(2r+3)$$

Po dosadení do (21) a vykrátení druhý sčítanec v (21) nadobudne tvar

$$-15\,i^2/(2r-1)(2r+1)(2r+3)$$

čo je tvar požadovaný v tvrdení vety. Veta je dokázaná.

Dôsledok. Pretože tvar (6) a (13) sú identické, polynomické kĺzavé priemery stupňa 2 a 3 sú jediné kĺzavé priemery, ktoré minimalizujú index redukcie chyby.

Literatúra

[1] Benjamin, B., Pollard, J.H.: *The Analysis of Mortality and Other Actuarial Statistics.* 2nd ed. Butterworth - Heinemann, Oxford 1980.

[2] Benjamin, B., Pollard, J.H.: *The Analysis of Mortality and Other Actuarial Statistics.* The Institut of Actuaries and the Faculty of Actuaries, London 1993.

[3] Cipra, T.: *Analýza časových řad s aplikacemi v ekonomii*. SNTL/Alfa, Praha 1986.

[4] Henderson, R.: *Mathematical Theory of Graduation*. Actuarial Society of America, New York 1938.

[5] Hindls, R., Kaňoková, J., Novák, I.: *Metody statistické analýzy pro ekonomy*. Management Press, Praha 1997.

[6] Hruška, Ľ.: *Identifikácia trendovej zložky časového radu pomocou kĺzavých priemerov*. Diplomová práca. FMFI UK Bratislava, 2006.

[7] Kendall, M.G., Stuart, A.: *The Advanced Theory of Statistics. Vol. 3. Design and Analysis, and Time-series.* Griffin, London 1966.

[8] Mészáros, J.: *Výpočet úmrtnostných tabuliek. Výpočet stratených rokov života*. Metodický materiál. Infostat, Bratislava 2000.

[9] Pastor, K.: *Graduačné formuly v programe STATGRAPHICS*. In: 7. Seminár Výpočtová štatistika, Bratislava 3.-4.12.1998. Zborník príspevkov. SŠDS Bratislava 1998, s. 72-76.

[10] Rublíková, E., Artl, J., Artlová, M., Libičová, L.: *Analýza časových radov*. Ekonóm, Bratislava 2003.

[11] Švecová, D.: *Metódy konštrukcie úmrtnostných tabuliek*. Diplomová práca. MFF UK Bratislava, 1999.

Riešené v rámci grantovej úlohy 1/3016/06.

Adresa autora:
Doc. RNDr. K.Pastor, CSc. Katedra aplikovanej matematiky a štatistiky FMFI UK,
Mlynská dolina, 842 48 Bratislava.
e-mail: pastor@fmph.uniba.sk

ÚLOHOU ŠTATISTIKY JE UKÁZAŤ SVET TAKÝ, AKÝ JE![1]

Jozef Chajdiak

Abstract: The paper contains some ideas about problems of the statistics.

Štatistika je súčasťou procesov rozhodovania. Nerozhoduje. Jej úlohou je len **ukázať svet taký, aký je**.

Jeden z rozmerov sveta je **čas**. Asi začína v mínus nekonečne a končí v plus nekonečne. Čas asi plynie len dopredu. Zdá sa, že tempo (rýchlosť) času v určitých podmienkach môžeme považovať za rovnomerné, vo všeobecnosti sa zdá, že čas plynie nerovnomerne. Jednou z čiastkových úloh štatistiky je ukázať svet taký, aký je v čase, ukázať jeho vývoj.

Klasické chápanie sveta je trojrozmerné. Naše tradičné chápanie priestoru je v **Euklidovskej metrike**. Poznáme a používame aj iné metriky. Možno sú ešte aj ďalšie metriky. Jednou z čiastkových úloh štatistiky je ukázať svet taký, aký je v niektorej z metrík.

Svet je **štrukturovaný**. Skladá sa z častí (objektov), vymedzených v priestore a čase. Pohľad na svet môžeme zdrobňovať v limite až po nulové zmeny času a nulové rozmery. Naopak, menšie časti môžeme agregovať do väčších, v limite až do jedného celku. V určitom časovom intervale a v určitom priestore (hrúbka) môžeme špecifikovať relatívne ucelené väčšie časti skladajúce sa z množiny menších častí (súbor ako množina jednotiek). Jednou z čiastkových úloh štatistiky je definovať štatistický súbor ako množinu štatistických jednotiek.

Časti sveta sú relatívne samostatné, sú navzájom prepojené, sú súčasťou jedného sveta (medzi rečou - nie je viac svetov?).

Svet sa priebežne mení. Asi platí, že na každú **akciu** nasleduje reakcia. Existuje veľké množstvo akcií a tomu zodpovedajúcich reakcií. Akcie vyjadrujú stav a zmenu časti sveta, reakcie dokončenie zmeny stavu časti sveta a jej nový stav. Tak ako máme malé/veľké časti sveta, tak máme malé/veľké akcie a im zodpovedajúce reakcie, tak máme malé/veľké stavy a zmeny v stavoch sveta. Úlohou štatistiky je merať tento vývoj, merať stav a zmeny v čase jednotlivých častí sveta a aj sveta ako celku.

Časť reakcií môžeme predpovedať pomerne spoľahlivo aj čo sa týka času, aj čo sa týka intenzity, aj čo sa týka miesta. Častejší je však prípad, že v podmienkach reakcie je prítomná aj určitá miera neurčitosti – reakcie majú **náhodný prejav**. Skúmanie charakteru náhodného prejavu (typ prejavu a jeho parametre resp. zákon rozdelenia pravdepodobnosti s jeho parametrami) je dôležitou úlohou štatistiky. Otázkou je, či náhodný je len prejav a objektívna podstata v pozadí má deterministický charakter a keďže stupeň poznania je ohraničený, tak ho vnímame ako náhodný, alebo aj v podstate analyzovaných javov či procesov je prítomná náhoda. V každom prípade v analyzovaných prejavoch náhodnosť

[1] Príspevok bol spracovaný v rámci riešenia grantovej úlohy VEGA 1/2631/05 „Analýza možností aplikácie viacrozmerných štatistických metód na skúmanie ekonomických výsledkov na príklade priemyslu SR prípadne iných oblastí ekonomiky“.

existuje a v rámci teórie máme viacero zákonov rozdelenia pravdepodobnosti a v rámci praktických poznaní viacero radov rozdelenia početnosti a v prípade analýzy v rámci chodu času niekoľko typov stochastických procesov s ich parametrami. Úlohou matematickej štatistiky a pravdepodobnosti je **formalizovať** tieto **zákony rozdelenia pravdepodobnosti** a **stochastické procesy**, úlohou štatistiky je po zmeraní stavu a vývoja analyzovanej časti sveta **odhadnúť**, podľa ktorého zákona rozdelenia pravdepodobnosti alebo stochastického procesu sa analyzovaný jav či proces riadi.

Svet (objektívna realita) určite nejaký **je**. Problémom je, ako ho vidíme. Úlohou štatistiky je ukázať svet taký, aký je. Znamená to opísať (popísať) príslušnú časť sveta, ktorá nás zaujíma, množinou ukazovateľov, definovať obsah týchto ukazovateľov a ich meracie jednotky, spôsob odmerania (zmerania, nameranie) hodnôt týchto ukazovateľov alebo spôsob ich odhadu, špecifikovať príslušnú časť sveta ako jeden celok a ak treba, aj ako časť sveta skladajúcu sa z množiny v nejakom zmysle podobných podčastí (súbor štatistických jednotiek), všetko v určenom časom momente. Ďalším aspektom pohľadu na svet je vidieť vývoj tejto časti sveta (len ako celku alebo len ako množiny jednotiek) reprezentovanej množinou ukazovateľov, meracích jednotiek, spôsobom merania resp. odhadov hodnôt v čase.

V analytických úlohách nastupuje problém, či ukazovatele (premenné) majú **spojitý** alebo **diskrétny** charakter. S diskrétnymi premennými je situácia pomerne jasná. Problémom môže byť, že reálne sme v stave zistiť (namerať) len diskrétne hodnoty (problém „šíp stojí“, hoci na druhej strane možno zistiť hodnoty až s takou presnosťou, že sa nám zdajú spojité) a v matematike, t. j. v matematickej štatistike a v pravdepodobnosti často pracujeme so spojitými premennými a ich analýzou.

Na príslušnú časť sveta môžeme pozerať ako na jeden celok alebo na množinu podčastí, z ktorých sa skladá (súbor jednotiek). V tejto súvislosti dôležitým problémom je problém **agregácie** výsledkov za časti do jedného celku reprezentujúceho príslušnú časť sveta ako jeden celok. Vzťahy medzi hodnotami podčastí môžu byť multiplikatívne, môžu byť aditívne, môžu byť zmiešané, a môžu byť aj iného charakteru. Problémom agregácie je, či súčin resp. súčet za všetky podčasti (jednotky) dáva celok (v ekonomickej štatistike agregácia typu hrubého výrobného obratu) alebo tento agregovaný výsledok treba ešte nejakým spôsobom korigovať (v ekonomickej štatistike agregácia typu finálneho produktu).

Naše predstavy o svete sú **modelom** objektívnej situácie, ktorú chceme (alebo žijeme v presvedčení, že chceme, hoci analyzujeme niečo iné) analyzovať. Ako každý model, aj naše predstavy sú zjednodušeným obrazom objektívnej skutočnosti. To spôsobuje, že naše predstavy o objektívnej realite môžu byť v určitej miere neúplné, môžu byť skreslené, prípadne od časti podstaty objektívnej reality viac či menej abstrahujeme.

V zásade existujú tri spôsoby štatistického skúmania, ktoré predstavujú tri rozličné modely. Podstatou prvého je **úplné zisťovanie** hodnôt a na ich analýze získanie predstavy o svete, takom akým je. Aj keď úplné zisťovanie dáva svojim spôsobom najširšie pohľady pre určenie obrazu sveta, problémom môžu byť potenciálne chyby pri úplnom zisťovaní (viac analyzovaných jednotiek, viac zisťovateľov a tým väčšia možnosť prítomnosti aj menej kvalifikovaného resp. spoľahlivého zisťovateľa, tým väčšia možnosť výskytu nepresne zistených hodnôt, tým nižšia miera spoľahlivosti výsledného obrazu).

Podstatou druhého spôsobu je **neúplné zisťovanie**. Keďže štatistické súbory sú zvyčajne rozsiahlejšie, zisťovanie hodnôt za každú jednotku môže byť pracné, nákladné a zdĺhavé. Riešením problému nedostatku chute, času alebo peňazí je neúplné zisťovanie.

Odborne seriózné neúplné zisťovanie (výberové skúmanie) je také, že každá jednotka súboru sa má možnosť dostať do vzorky (výberu) a pre každú jednotku poznáme aj pravdepodobnosť jej zaradenia do vzorky (výberu). Na základe zistených hodnôt odhadujeme obraz sveta takého aký je. Problémom neúplného zisťovania je skupina postupov, pri ktorých buď nie každá jednotka môže byť zahrnutá do vzorky (výberu) alebo častejšie – nepoznáme pravdepodobnosti zaradenie do vzorky (výberu) pre všetky jednotky súboru. Odhady sveta takého aký je na základe takto získaných hodnôt za vybrané jednotky sú lepšie ako nič a možno aj postačujú ako podklad k rozhodovaniu, ale do vedeckej serióznosti tu predsa len niečo chýba. Najčastejšie problémy sú pri dotazníkových prieskumoch, keď poznáme pravdepodobnosť zaslania dotazníka každej jednotke, ale, napríklad pri 30 % návratnosti, nevieme určiť pravdepodobnosť návratu vyplneného dotazníka od príslušnej jednotky. Použitie modelu jednoduchého náhodného výberu či iného modelu v takejto situácii môže byť aj silne nekorektné.

Podstatou tretieho spôsobu je **experimentálne zisťovanie**. Postatou problému pri experimentálnom zisťovaní je ani nie tak presne ukázať svet taký, aký je ale, ukázať príslušnú časť sveta tak, aby bola taká ako chceme a potom kontrolovať, či je taká ako chceme (t.j. prakticky cez ukazovanie sveta taký aký je, kontrolovať, či je taký aký sme chceli aby bol).

Kvalita získaných údajov predurčuje kvalitu obrazu sveta, takého aký je. Kvalitným údajom je taký údaj, ktorý v dostatočnej miere odráža objektívnu realitu. Každý vykázaný **údaj** (údaj, ktorý analyzujeme) sa potenciálne skladá zo štyroch zložiek. Základom údaju je obsah toho, čo analyzujeme, je odraz objektívnej reality. K tejto objektívnej podstatne sa pridružujú tri typy chýb: náhodná chyba, systematická chyba a chyba z titulu nepoznania. Náhodné chyby majú zvyčajne normálne rozdelenie so strednou hodnotou nula (t.j. chyby sa navzájom kompenzujú a výskyt väčších chýb je menej pravdepodobný). Systematické chyby predstavujú vplyv faktora, ktorý ovplyvňuje namerané výsledky v sebe želanom smere, t. j. určitým systematickým smerom odchýlne od objektívnej reality. Systematické chyby skresľujú výsledok jedným smerom a trend ich vývoja je zvyčajne paralerny s trendom vývoja objektívnej reality. Chyby z titulu nepoznania (resp. neúplnej, či chybnej špecifikácie ukazovateľa a spôsobu jeho zisťovania) sú potenciálne možné a zvyčajne žijeme v presvedčení, že neexistujú.

V prípade, že máme kvalitné údaje, je možné z nich dostať kvalitné výsledky analýz, dostatočne presný obraz sveta, odraz sveta takého aký je. Samozrejme nekorektné postupy analýz dávajú výsledky primerané miere ich nekorektnosti.

Z nekvalitných údajov môžeme dostať len nekvalitné výsledky analýz, nekvalitný obraz sveta, takého aký je. Najčastejším spôsobom nekvality analyzovaného súboru údajov je úprava alebo vylúčenie neželaných extrémnych napozorovaných hodnôt (mení sa tým variabilita v súbore), v prípade výberov neúplná realizácia výberového zisťovania (získanie údajov len za časť jednotiek, za ktoré sme mali zistiť príslušné hodnoty).

V svete a jeho častiach nastávajú rozličné javy, prebiehajú rozličné procesy, existujú rozličné subjekty a objekty. Vnímame, pozorujeme **prejavy** týchto javov, procesov, subjektov a objektov. Problémom je, do akej miery je zhodné naše poznanie prejavu príslušného javu, procesu, subjektu alebo objektu so skutočnosťou.

Medzi jednotlivými javmi existujú rozličné **väzby**. Úlohou štatistiky je vyhodnotiť charakter a intenzitu týchto väzieb. Podobne procesy majú svoje parametre. Aké sú to parametre, aké sú to hodnoty? Aké sú väzby medzi parametrami jednotlivých procesov?

Mieru nášho poznania sveta vyjadruje **konkrétna historická skúsenosť** poznávajúceho. Konfrontuje získané vedomosti so skutočnosťou, konfrontuje svoje poznatky so skutočnosťou, konfrontuje, či obraz stavu a vývoja skutočnosti sa zhoduje so stavom a vývojom skutočnosti.

Štatistika nie je ľahká disciplína. **Jej úlohou je ukázať svet taký, aký je**!

Kontakt na autora:
Doc. Ing. Jozef Chajdiak, CSc.
chajdiak@statis.biz

Zamestnanosť podľa odvetví ekonomickej činnosti v štátoch EÚ

Beáta Stehlíková

Abstract
In the theoretical literature, the mechanisms promoting spatial concentration of the employment have been described. The aim of the paper is the analysis of the employment in sectors in European union using statistical methods.

Dôležitým determinantom, ktorý môže konvergenčný proces urýchliť alebo spomaliť je spôsob, akým sa ekonomika Slovenska dokáže zapojiť do medzinárodnej deľby práce. Odvetvová špecializácia predstavuje stupeň, v akom sa zamestnanosť a pridaná hodnota sústreďujú v odvetviach ekonomiky jednej krajiny v porovnaní s druhou. Odvetvová špecializácia má vplyv na rast produktivity práce. Determinantmi rozloženia zamestnanosti v jednotlivých odvetviach sa zaoberajú viacerí autori, napríklad Haaland a kol. (1998), Puga (1999), Braunerhielm a kol. (2003). Z dlhodobého hľadiska sú štrukturálne zmeny, pri ktorých sa presúvajú zdroje z odvetví s nízkou produktivitou práce do produktívnejších impulzom rýchlejšieho rastu ekonomiky. Cieľom príspevku je štatistické posúdenie rozloženia zamestnanosti v jednotlivých odvetviach v štátoch Európskej únie.

Materiál a metódy

Odvetvia ekonomickej činnosti sú poľnohospodárstvo, poľovníctvo a lesníctvo (A), rybolov a chov rýb (B), ťažba nerastných surovín (C), priemyselná výroba (D), výroba a rozvod elektriny plynu a vody (E), stavebníctvo (F), veľkoobchod a maloobchod; oprava vozidiel (G), hotely a reštaurácie (H), doprava, skladovanie, pošty a telekomunikácie (I), finančné sprostredkovanie (J), nehnuteľnosti, prenájom a obchodné činnosti (K), školstvo (M), zdravotníctvo a sociálna pomoc (N), verejná správa a obrana; povinné sociálne zabezpečenie (L), ostatné spoločenské, sociálne a osobné služby (O), činnosti domácností (P), exteritoriálne organizácie (Q). V analýzach sú kategórie A, B a L, O, P, Q spojené. Údaje o zamestnanosti podľa odvetví ekonomickej činnosti (OKEČ) v štátoch Európskej únie za rok 2004 sme čerpali z databázy ILO.
Koncentrácia zamestnanosti podľa OKEČ bola posúdená pomocou polohového koeficienta. Polohový koeficient umožňuje posúdiť koncentráciu zamestnanosti v určitej kategórii v danom štáte v porovnaní s EÚ ako celkom. Vypočítame ho podľa vzťahu

$$LQ_i = \frac{A_i / \sum A_i}{B_i / \sum B_i},$$

kde n je počet štátov (25), A_i je intenzita zamestnanosti v určitej kategórii A v štáte i (i = 1, 2, …, n), B_i je celková zamestnanosť v štáte i (i = 1, 2, …, n). Ak je hodnota polohového koeficientu väčšia ako 1 (menšia ako 1), indikuje to koncentráciu zamestnanosti v určitej kategórii v štáte i väčšiu (menšiu) v porovnaní s EÚ ako celkom.

Giniho index (index rozdielnosti)

$$G = \sum \left| A_i / \sum A_i - B_i / \sum B_i \right| / 2$$

je mierou relatívnej koncentrácie zamestnanosti v určitej kategórii vo vzťahu k báze – celkovej zamestnanosti. Nadobúda hodnoty z intevalu od 0 do 100. Ak sa rovná nule, potom koncentrácia zamestnanosti v určitej kategórii sa rozprestiera v súlade s celkovou zamestnanosťou. Ak sa rovná hodnote sto, potom je zamestnanosť v určitej kategórii koncentrovaná do jedinej oblasti.

Zo štatistického hľadiska je polohový koeficient podiel, t.j. $LQ_i = \frac{\alpha}{\beta} = \theta$. Kovariančná matica odhadov $(\hat{\alpha}, \hat{\beta})$ je matica $\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$. Platí $V_{11} = \frac{\frac{A_i}{B_i}\left(1 - \frac{A_i}{B_i}\right)}{B_i}$, $V_{22} = \frac{\sum \frac{A_i}{B_i}\left(1 - \frac{A_i}{B_i}\right) B_i}{(\sum B_i)^2}$, $V_{12} = V_{21} = \frac{\frac{A_i}{B_i}\left(1 - \frac{A_i}{B_i}\right)}{\sum B_i}$. Rozptyl odhadu $\hat{\theta}$ je $\left(V_{11} - 2\hat{\theta}V_{12} + \hat{\theta}^2 V_{22}\right)\hat{\beta}^{-2}$. Interval spoľahlivosti, vypočítaný pomocou delta metódy, pre normálne rozdelenie, je

$$\left(\hat{\theta} - z_{1-\alpha/2}\hat{\sigma}, \hat{\theta} + z_{1-\alpha/2}\hat{\sigma}\right).$$

Normalitu rozdelenia vzhľadom na nízku hodnotu n = 25, sme overili pomocou Shapirovho-Wilkovho testu.

Podobné štáty z hľadiska vysokých technológií sme určili pomocou zhlukovej analýzy. Zo zhlukovacích metód bola zvolená Wardova metóda, ktorá minimalizuje vnútrozhlukový rozptyl. Vzdialenosť sme určili pomocou klasickej Euklidovej vzdialenosti.

Výpočty boli realizované pomocou softvéru SAS.

Výsledky a diskusia

Najvyššiu hodnotu nadobúda Giniho index pre ťažbu nerastných surovín – 52,50 percent. Druhú najvyššiu hodnotu 23,87 nadobúda zamestnanosť v poľnohospodárstve, poľovníctve, lesníctve, rybolove a chove rýb. Tretie miesto v koncentrácii patrí zamestnanosti v odvetví výroby a rozvodu elektriny plynu a vody – 20,59. Giniho indexy pre zamestnanosť v ostatných odvetviach sú H - 18,62, K - 14,50, N - 12,18, D - 12,05, F - 11,28, J - 10,62. V súlade s celkovou zamestnanosťou je zamestnanosť v školstve – 9,09, zamestnanosť v spojených odvetviach LOPQX – 7,20, zamestnanosť vo veľkoobchode, maloobchode a oprave vozidiel – 5,79. Giniho index nadobúda najnižšiu hodnotu pre zamestnanosť v doprave, skladovaní, poštách a telekomunikáciách – 5,22.Variačné rozpätie Giniho koeficienta indikuje veľkú rozdielnosť v zamestnanosti v jednotlivých odvetviach – od takmer rovnomerného rozloženia (5,22) až do hodnoty 52,50.

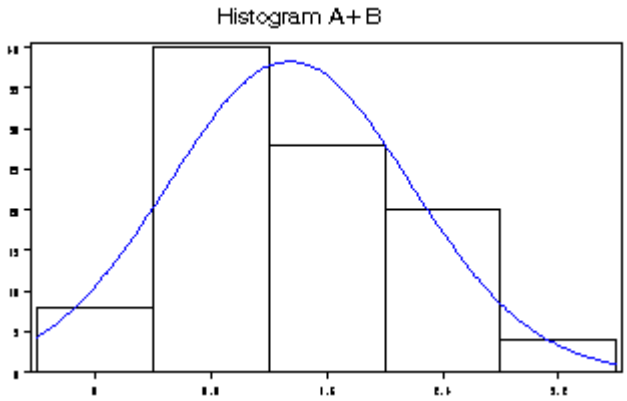
Výsledok Shapirovho Wilkovho testu normality pre malý rozsah súboru (25 štátov Európskej únie) je v tabuľke 1. Normálne rozdelenie majú podiely zamestnania v kategórii poľnohospodárstvo, poľovníctvo a lesníctvo, rybolov a chov rýb (AB), priemyselná výroba (D), stavebníctvo (F), doprava, skladovanie, pošty a telekomunikácie (I), školstvo (M) a v spojenej kategórii LOPQX.

Tabuľka 1 Výsledok Shapirovho Wilkovho testu normality

OKEČ	Hodnota testovacej štatistiky	P hodnota
A+B	0,940857	0,1549
C	0,289546	0,0001
D	0,960326	0,4210
E	0,774694	0,0001
F	0,905708	0,0245
G	0,857127	0,0024
H	0,877893	0,0063
I	0,951355	0,2689
J	0,680479	0,0001
K	0,864498	0,0034
M	0,934833	0,1124
N	0,869615	0,0040
LOPQX	0,945884	0,2022

Zdroj: Vlastné výpočty

Obrázok 1 Histogram pre zamestnanosť v poľnohospodárstve, poľovníctve, lesníctve, rybolove a chov rýb



Zdroj: Vlastné výpočty

V tabuľke 2 sú intervaly spoľalivosti pre koeficienty lokalizácie preukazne vyššie ako 1 pre OKEČ AB, D, F, I, M, LOPQX. Interval spoľahlivosti kvôli nenormálnemu rozdeleniu príslušných koeficientov lokalizácie nebolo možné korektne pomocou delta metódy skonštruovať. Ďalší možný prístup nevyžadujúci normalitu, predstavujú zovšeobecnené lineárne modely. V tomto prípade však dávajú veľmi široké konfidenčné intervaly.

Tabuľka 2 Intervaly spoľahlivosti pre koeficienty lokalizácie

Štát	AB		D		F		I		M		LOPQ	
Belgicko	0,19	0,46	0,90	1,03	0,72	0,94	1,19	1,48	1,18	1,42	1,03	1,20
Cyprus	1,25	3,50	0,39	0,73	1,18	2,16	0,47	1,24	0,47	1,11	0,89	1,45
ČR	1,56	2,13	1,46	1,61	0,89	1,11	1,15	1,42	0,79	0,99	0,76	0,90
Dánsko	0,58	1,07	0,77	0,92	0,76	1,03	0,95	1,26	0,92	1,19	0,75	0,93
Estónsko	1,09	2,68	1,07	1,44	0,75	1,38	1,02	1,78	0,94	1,57	0,61	0,99
Fínsko	0,62	1,19	0,03	0,07	1,53	1,92	0,27	0,47	1,82	2,22	0,54	0,71
Francúzsko	0,64	0,79	0,88	0,92	0,82	0,91	1,05	1,15	0,91	0,99	1,22	1,28
Grécko	0,44	0,89	0,73	0,87	1,13	1,45	1,04	1,37	1,10	1,39	1,21	1,42
Holandsko	1,25	1,60	0,65	0,73	0,80	0,95	0,84	1,00	0,74	0,87	0,93	1,03
Írsko	0,42	1,02	0,77	0,96	1,21	1,64	0,77	1,15	0,77	1,10	0,71	0,93
Lotyšsko	2,58	4,21	0,76	1,01	1,00	1,53	1,34	1,98	0,92	1,39	0,82	1,14
Litva	1,53	2,67	0,94	1,17	1,10	1,57	0,96	1,44	1,26	1,73	0,65	0,90
Luxembursko	-0,17	0,55	0,38	0,76	0,94	1,95	0,88	1,93	0,28	0,90	0,57	1,10
Madarsko	1,45	2,12	1,22	1,38	0,60	0,83	1,11	1,43	1,00	1,27	0,91	1,09
Malta	-0,35	1,48	0,70	1,41	0,29	1,50	0,48	1,96	0,58	1,87	0,57	1,42
Nemecko	0,58	0,70	1,22	1,26	0,88	0,95	0,89	0,96	0,73	0,80	0,99	1,04
Poľsko	0,92	1,19	1,22	1,30	0,83	0,96	1,03	1,18	1,21	1,35	0,82	0,90
Portugalsko	1,04	1,54	1,12	1,26	1,53	1,82	0,68	0,90	0,82	1,03	1,00	1,17
Rakúsko	2,14	2,82	0,87	1,00	1,04	1,29	0,89	1,14	0,61	0,79	0,78	0,93
SR	2,00	2,95	1,34	1,55	0,92	1,25	0,89	1,25	0,88	1,19	0,72	0,93
Slovinsko	1,43	2,79	1,40	1,72	0,56	1,01	0,69	1,21	0,74	1,20	0,71	1,02
Španielsko	1,48	1,74	0,90	0,96	1,70	1,84	0,87	0,98	0,77	0,86	1,05	1,13
Švédsko	0,31	0,60	0,79	0,91	0,63	0,83	0,90	1,14	1,40	1,66	0,71	0,85
Taliansko	1,27	1,50	1,26	1,32	0,95	1,06	0,89	1,00	1,09	1,20	1,14	1,22
UK	0,37	0,48	0,61	0,65	0,66	0,73	0,90	0,99	1,06	1,14	0,75	0,80

Zdroj: Vlastné výpočty na základe údajov ILO

Zamestnanosť v poľnohospodárstve je na Cypre, v ČR, Estónsku,Holandsku, Litve, Lotyšsku, Maďarsku, Portugalsku, Rakúsku, Slovensku, Slovinsku, Španielsku a Taliansku vyššia ako v EÚ ako celku. Koncentrácia zamestnanosti v školstve je vyšia ako v EÚ v Belgicku, Fínsku, Grécku, Litve, Maďarsku, Poľsku, Švédsku, Taliansku a UK. Zamestnanosť v priemysle v ČR, Estónsku, Maďarsku, Nemecku, Poľsku, Portugalsku, SR, Slovinsku, Taliansku je vyššia ako v EÚ.

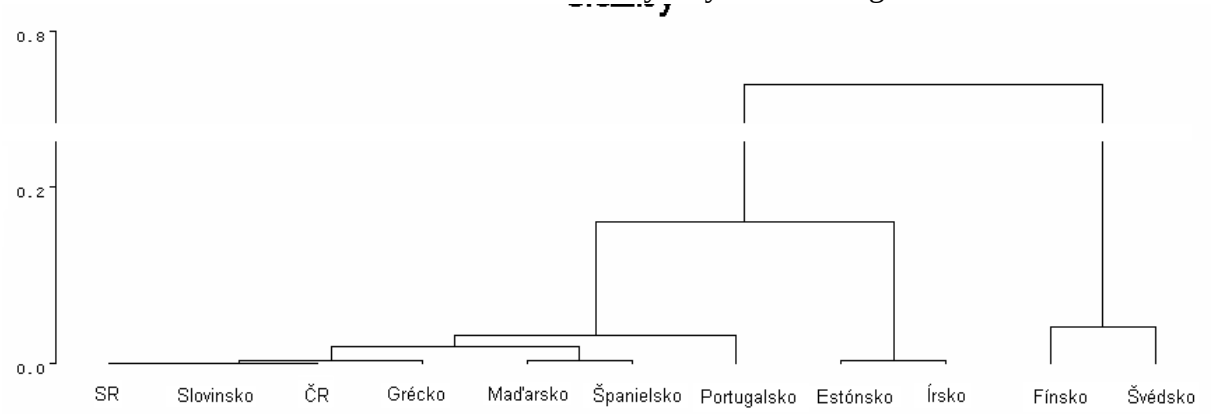
V príspevku sa nezaoberáme druhou zložkou odvetvovej špecializácie - pridanou hodnotou jednotlivých odvetví slovenskej ekonomiky. Čutková, Donoval (2004) konštatujú, že relatívne vyšší podiel pridanej hodnoty sa v SR v porovnaní s priemernou štruktúrou ekonomík Európskej únie vytvorí v priemysle a pôdohospodárstve. Je to v súlade so štatisticky preukazne vyššou koncentráciou zamestnanosti v týchto odvetviach. Ďalej konštatujú, že podiel služieb na tvorbe pridanej hodnoty v EÚ je v priemere 70 %, kým v SR služby prispievajú k celkovej tvorbe pridanej hodnoty 60 %.

Šikulová (2004) uvádza, že krajiny produkujúce technologicky a poznatkovo náročné tovary a služby sú schopné zvyšovať svoj podiel na rastúcich a náročných trhoch a prostredníctvom vyšších cien produkcie dosahovať vyššie dôchodky v relácii k jednotlivým výrobným faktorom.

V rámci spracovateľského priemyslu sa za stredne vysoké technológie sa považujú chemikálie a vlákna, všeobecné strojárstvo, elektrické stroje a prístroje, motorové vozidlá, ostatné dopravné prostriedky. Za vysoké technológie sa považujú kancelárske stroje, rádiové, TV a komunikačné zariadenia a prístroje, zdravotnícke, presné a optické prístroje. Poznatkovo náročné služby v sektore služieb sa považujú pošta a telekomunikácie, počítačové a súvisiace činnosti, výskum a vývoj. Poznatkovo náročné trhové služby sú vodná a letecká doprava, činnosti v oblasti nehnuteľnosti, prenájom strojov a zariadení, iné obchodné služby. Výsledok zhukovania za štáty, za ktoré boli k dispozícií údaje sú na obrázku 2 a 3.

Slovenská republika nedosahuje z hľadiska zamestnanosti parametre EÚ ani v jednej zo skupín sektora služieb. Podstatne nižší je najmä podiel zamestnanosti v poznatkovo náročných trhových službách. Významným faktorom úspešnosti podnikov je efektívne poskytovanie a využívanie poznatkovo náročných služieb. Medzera medzi podielom zamestnanosti v sektore poznatkovo náročných službách u nás a v štátoch EÚ je prejavom celkove nižšej technologickej úrovne našej ekonomiky. Pri nízkej inovačnej aktivite podnikov SR, nízkych výdavkoch na výskum a vývoj, malej podpore výskumných a inovačných aktivít zo strany štátu, chýba impulz na intenzívnejší rozvoj poznatkovo náročných služieb. Ďalším dôvodom je skutočnosť, že firmy z vyspelých ekonomík presúvali do krajín s nižšou mzdovou úrovňou predovšetkým tie fázy produkcie, ktoré sú náročné na ľudskú prácu. Veľká časť poznatkovo náročných služieb súvisiacich s produkciou zostáva pritom v materskej krajine.

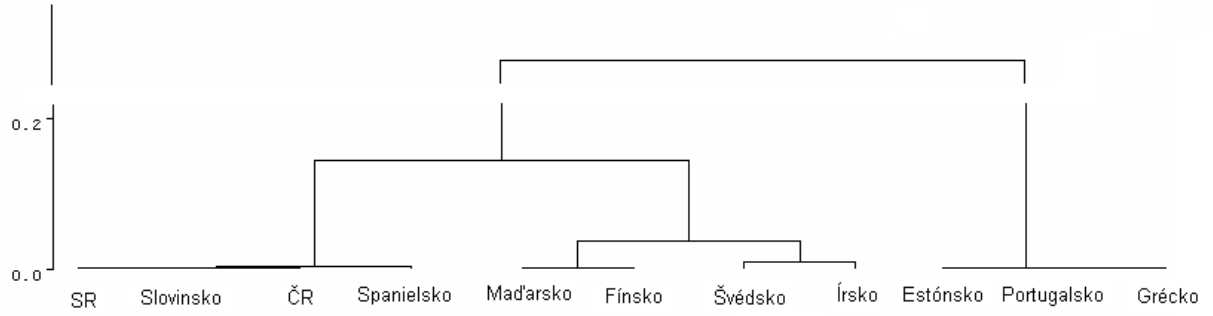
Obrázok 2 Zamestnanosť v službách z hľadiska vysokých technológií



Zdroj: Vlastné výpočty

Rozdiely medzi EÚ(15) a SR sa v odvetviach spracovateľského priemyslu s vysokými a stredne vysokými technológiami sú pre SR priaznivé – podiel zamestnaných v týchto odvetviach dokonca mierne prevyšuje priemernú úroveň EÚ (15).

Obrázok 3 Zamestnanosť v spracovateľskom priemysle z hľadiska vysokých technológií



Zdroj: Vlastné výpočty

Použité štatistické metódy naznačili odlišnosti v štruktúre zamestnanosti v jednotlivých štátoch Európskej únie ako aj odlišnosti v štruktúre zamestnanosti Slovenska a zamestnanosti v Európskej únii ako celku. Základnou črtou rozdielov zamestnanosti Slovenska a zamestnanosti v Európskej únii je menej rozvinutý sektor služieb na jednej strane a vyššia viazanosť pracovných síl v spracovateľskom priemysle a v ostatných odvetviach. V budúcnosti bude potrebné analyzovať ako integrácia Slovenska do EÚ ovplyvnila zamestnanosť v jednotlivých odvetviach a jeho odvetvovú špecializáciu. Odvetvová špecializácia vplýva na rast ekonomiky z krátkodobého aj dlhodobého hľadiska. Preto bude potrebné, aby sa do procesu odvetvovej špecializácie razantnejším a kvalifikovanejším spôsobom vstupovali všetky odvetvové rezorty vlády SR.

Literatúra

Šikulová I.: Štrukturálna konvergencia Slovenskej republiky k EÚ 15 – predpoklad úspešného členstva Slovenska v európskej menovej únii. Ekonomický časopis, 52, č. 9, 2004. str. 1064 – 1079.

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

Thacker, N. A. , Aherne, F. J. , Rockett, P. I.: The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data. Kybernetika, 34, č. 4, 1997. str. 363-368

Haaland, J., Kind, H., Midelfart-Knarvik, K., Torstenson, J.: What determines the economic geography of Europe? Discusion paper 19/98 Bergen: EC, 1998

Čutková, Z., Donoval, M.: Odvetvová špecializácia v SR. BIATEC, 12, č.8, 2004. str. 6-9

Puga, D.: The rise and fall of regional inequalities. European economic review. 43,1999. str. 303-335

Braunerhielm,P., Johanson,D: Determinants of spatial concentration. Journal of Industry Studies. 2003 Dokument vo formáte HTML
http://www.findarticles.com/p/articles/mi_qa3913 [cit. 2006-04-04]

European Trend Chart on Innovation. www.cordis.lu/trendchart [cit. 2006-04-04]

European Innovation Scoreboard, 2003. www.cordis.lu/trendchart [cit. 2006-04-04]

Databáza ILO. http://laborsta.ilo.org/ [cit. 2006-04-04]

Kontaktná adresa

doc. RNDr. Beáta Stehlíková, CSc., Katedra štatistiky a operačného výskumu, FEM SPU Nitra. Tr. A. Hlinku 2, 949 76 Nitra, Slovenská republika
e-mail:Beata.Stehlikova@uniag.sk

ZÁKLADNÉ PRINCÍPY STRATÉGIE MODELOVANIA

Iveta Stankovičová

Abstract

There are typically two goals of mathematical modeling. One is to obtain a valid estimate of relationship between variables and the other is to obtain a good predictive model. The modeling strategy involves three stages: 1) variable specification, 2) interaction assessment, and 3) confounding assessment followed by consideration of precision. The initial model has to be hierarchically well-formulated and illustrated.

Úvod

ri hľadaní „najlepšieho“ viacnásobného lineárneho alebo logistického modelu s interakčnými premennými sa odporúča uplatňovať stratégiu, ktorá sa skladá z nasledovných troch krokov:

1. špecifikácia premenných,
2. vyhodnotenie interakcií medzi premennými a
3. posúdenie modifikácie efektov modelu ako následok precíznosti modelu.

Počiatočný východiskový (inicializačný) model musí byť hierarchicky správne zostavený a vysvetliteľný. Na získanie počiatočného modelu sa odporúča použiť hierarchickú eliminačnú metódu „backward“, ktorá umožní zo všetkých uvažovaných premenných vybrať len tie, ktoré sú štatisticky významné vo vzťahu k modelovanej závislej premennej na zvolenej hladine významnosti. Štatistické testy umožňujú zhodnotiť aj význam interakčných výrazov v modeli, ale neumožňujú posúdiť nepresnosť záverov z interakcií.

Princípy stratégie modelovania

V mnohých výskumných štúdiách sa používajú matematické modely (napr. viacnásobná logistická regresia) a pri prezentovaní dosiahnutých výsledkov v odborných článkoch autori poskytujú len málo informácií o použitých metódach a modelovacej stratégii. Často neposkytujú informácie ako boli vybraté premenné pre počiatočný model, ako (akou metódou) boli selektované premenné pre konečný model, ako boli modifikované efekty a vyhodnotené chyby. Bez informácií o použitej stratégii modelovania, je ťažké ohodnotiť spoľahlivosť (validitu) výsledného modelu.

V praxi sa pri modelovaní často používa stratégia *„ad hoc“*, čiže sa počas analýzy improvizuje. Zásada je, aby všetci výskumníci boli v stratégii jednotný.

V zásade sa v matematickom modelovaní vyskytujú dva ciele (zámery):

1. odhad parametrov modelu alebo
2. výpočet predikcie (resp. predpovede).

Od týchto cieľov závisí aj stratégia použitá pri modelovaní. Keď je cieľom získať predikciu, tak je vhodné použiť štandardné výpočtové algoritmy, ako sú napríklad eliminačné postupy backward, fordward a stepwise, ktoré sú súčasťou každého štatistického softvéru. Ak je cieľom modelovania získať spoľahlivé odhady parametrov modelu, čiže efekty, tak tento

prístup nie je vhodný. V tomto prípade je potrebné si dávať pozor na premenné, ktoré majú vplyv na parametre modelu a ovplyvňujú ich hodnoty, respektíve ich modifikujú. Práve v tomto prípade sa odporúča použiť stratégiu modelovania skladajúcu sa z troch krokov, spomínaných v úvode tohto príspevku.

Špecifikácia premenných je prvým krokom modelovania, v ktorom výskumník zaraďuje do modelu všetky zmysluplné premenné a vychádza pritom z literatúry, podobných výskumných štúdií a z vlastných poznatkov a hypotéz. V tomto kroku ide o vytvorenie inicializačného modelu s čo najväčším počtom premenných bez interakcií.

Vyhodnotenie interakcií medzi premennými je následným krokom v stratégii modelovania. Do modelu sa zaraďujú aj interakčné premenné rôzneho stupňa a hodnotí sa ich významnosť v modeli.

Posúdenie modifikácie efektov modelu ako následok precíznosti modelu nasleduje ako tretí krok stratégie modelovania. Ide o zistenie, či interakciu medzi dvoma premennými nemodifikuje alebo nevysvetľuje tretia premenná. Predpokladajme, že z odhadu parametrov modelu logistickej regresie sme zistili, že bodový odhad pomeru šancí (odds ratio = OD) pre interakčnú premennú AB je 3,5. Keď však odhadneme tento pomer šancí zvlášť podľa tretej premennej C, napr. nech je to pohlavie, zistíme že: OD pre mužov je 5,4 a OD pre ženy je len 1,2. Z výsledku je zrejmé, že interakcia premenných AB je rozdielna podľa pohlavia C a teda je dôsledkom pôsobenia C. Nie je správne interpretovať priemerné OD=3,5, lebo táto hodnota je zavádzajúca a reprezentuje len jednoduchý „priemer“ z pomerov šancí pre mužov a ženy.

Ďalej predpokladajme, že v hierarchickom logistickom modeli je významná interakcia troch premenných ABC. Bodový odhad pomeru šancí pre túto interakciu je OD = 2,4 a 95%-ný interval spoľahlivosti je (1,2; 3,7). V modeli je významná aj premenná A a jej pomer šancí OD = 6,2 a 95%-ný interval spoľahlivosti je (5,9; 6,4). Šírka intervalu spoľahlivosti pre pomer šancí interakcie je väčšia ako šírka intervalu pre samostatnú premennú A. Pri interpretácii modelu však uprednostníme pomer šancí OD = 2,4 pred vyšším OD = 6,2, lebo interakčnú premennú ABC pokladáme za dôležitejšiu vzhľadom k modelovanej premennej, ako samotnú premennú A. Výsledok 2,4 kontroluje naraz všetky relevantné premenné a tak poskytuje spoľahlivejšiu odpoveď ako výsledok 6,2, ktorý kontroluje len jednu premennú. Interval spoľahlivosti pre interakciu ABC je širší, čiže menej presný ale nevychýlený. Užší je interval spoľahlivosti pre premennú A a teda aj presnejší, ale je vychýlený.

V podstate v treťom kroku stratégie modelovania sa pokúšame nájsť taký odhad parametrov modelu, aby sme kontrolovali všetky vysvetľujúce premenné a naviac, aby model bol čo najprecíznejší.

Hierarchicky dobre zostavený model (hierarchically well-formulated model = HWF) má takú štruktúru, ktorá bráni vzniku zavádzajúcich výsledkov a môžeme ho charakterizovať takto: Keď je v modeli interakčná premenná vyššieho rádu, musia byť v modeli aj všetky premenné nižších rádov, z ktorých je interakcia zložená.

Príklad HWF modelu:

$$\text{logit}P(Y) = a + bA + cX_1 + dX_2 + \\ + eAX_1 + fAX_2$$

Prvky interakcie AX_1: A, X_1 sú v modeli
Prvky interakcie AX_2: A, X_2 sú v modeli

Príklad nie HWF modelu:

$$\text{logit}P(Y) = a + bA + cX_1 + dX_2 + \\ + eAX_1 + fAX_2 + gAX_1X_2$$

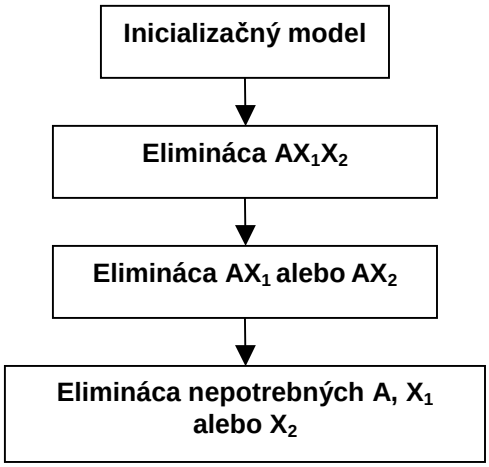
Prvky interakcie AX_1X_2: A, X_1, X_2, AX_1, AX_2 sú v modeli, ale X_1X_2 nie je v modeli

HWF model je vhodnejší pre testovanie významnosti interakčných premenných vyšších rádov ako inak zostavené modely. Predpokladajme, že premenná A je binárna a môže byť kódovaná rôznymi spôsobmi, napr. A=(0, 1) alebo A=(-1, 1). Testovanie významnosti premennej ABC je v HWF modeli nezávislé od spôsobu tohoto kódovania, kým v inak zostavených modeloch je závislé.

Keď sme si v etape špecifikácie zostavili rozsiahly inicializačný HWF model, odporúča sa na redukciu premenných a na získanie konečného modelu použiť ***hierarchickú spätnú (backward) eliminačnú metódu***. Toto riešenie zobrazuje nasledovný diagram:



Hierachická backward eliminácia



V hierarchickom spätnom (backward) eliminačnom procese sa na každej úrovni eliminácie ponechávajú významné a vylučujú nevýznamné premenné z modelu. Zachováva sa však ***hierarchický princíp***, ktorý je založený na tom, že ak je v modeli významná interakcia vyššieho rádu, tak sa v modeli zachovajú aj jej prvky, t.j. interakcie nižšie rádu. *Príklad:* Ak sú v modeli ponechané premenné AX_1 a AX_2 , tak model musí obsahovať aj premenné nižšieho rádu a to A, X_1 a X_2.

Ak je v HWF modeli významná interakcia vyššieho rádu, napr. AX_1X_2 tak jej test významnosti nezávisí na kódovaní binárnej premennej A, ale už jej nižšie prvky (AX_1 , AX_2 , X_1 a X_2) sú na tomto kódovaní závislé.

Záver

Dnes si nevieme predstaviť, že by sme robili štatistické analýzy bez vhodného softvéru. Na druhej strane, treba užívateľov vystríhať od používania počítačových procedúr a algoritmov bez dôkladného preštudovania si dokumentácie a literatúry, ktorá sa teoreticky zaoberá danou problematikou. Často sa stáva, že užívatelia nepoužívajú vhodné procedúry, alebo nesprávne interpretujú výsledky počítačových výstupov.

V modelovaní je potrebné použiť vhodnú stratégiu, ktorá je závislá od cieľa modelu. V modeloch viacnásobnej logistickej a lineárnej regresie s interakčnými výrazmi je potrebné pri eliminácii nevýznamných premenných uplatňovať hierarchický princíp: Presvedč sa, či máš HWF model. Ak je významná interakčná premenná AX_1X_2, tak preskúmaj, či zostali v modeli aj jej prvky nižšieho rádu (AX_1 , AX_2 , X_1 a X_2), ak nie, tak model nie je HWF.

Použitá literatúra

[1] Friendly Michael: Visualizing Categorical Data. SAS Institute Inc. Cary, 2000. ISBN 1-58025-660-0

[2] Hebák Petr, Hustopecký Jiří, Jarošová Eva, Pecáková Iva: Vícerozměrné statistické metody 1. INFORMATORIUM. Praha 2004. ISBN 80-7333-025-3.

[3] Kleinbuam, David G.: Logistic Regression. A Self-Learning Text. 1994 Springer–Verlag New York, Inc., ISBN 0-387-94142-8

[4] Stankovičová, Iveta: Grafy interakcií. Zobrazovanie výsledkov logistických modelov. Zborník príspevkov z 12. medzinárodného seminára Výpočtová štatistika. Bratislava. SŠDS 2003, str. 105-109. ISBN 80-88946-29-8

[5] Stankovičová, Iveta: Logická regresia a hĺbková analýza dát (Data Mining). Zborník príspevkov z 11. slovenskej štatistickej konferencie: Štatistické metódy v praxi. Bratislava, SŠDS 2002. ISBN 80-88946-19-0

[6] 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

Štatistické analýzy založené na redukcii výberového priestoru

Rastislav Potocký

Abstract: Dimension-reducing techniques are presented which condense information on objects under consideration thus making analysis of a data set easier.

ÚVOD

Informácia o sledovaných objektoch, ktorú sprostredkuje množina p štatistických znakov sa niekedy dá získať aj z menšieho počtu k (k< p) znakov. Ak je redukcia významná, uľahčuje to analýzu. Je zrejmé , že testy založené na jednorozmerných rozdeleniach majú lepšie vlastnosti ako viacrozmerné, podobne konštrukcia intervalov spoľahlivosti je jednoduchšia ako oblastí spoľahlivosti.

V tomto príspevku sa zameriame na štyri metódy :

a) metódu hlavných komponentov,

b) faktorovú analýzu,

c) kanonické korelácie ,

d) diskriminačné súradnice.

Ukážeme,že matematicky ide o príbuzné metódy, založené na minimalizácii nejakej vzdialenosti, a preto aj riešenia sú podobné : treba nájsť vlastné hodnoty a zodpovedajúce vlastné vektory vhodnej matice.

Najčastejšie používaná metóda spočíva v transformácii bodov v p-rozmernom priestore na body v k-rozmernom priestore, ktoré sú nejakým spôsobom blízke pôvodným bodom. Pritom k sa určí tak, aby zostala zachovaná podstatná časť celkovej disperzie st S, kde S je výberová kovariančná matica.To je princíp metódy hlavných komponentov.

Naopak faktorová analýza hľadá lineárny model s k novými premennými, ktorý vysvetľuje kovarianciu medzi pôvodnými štatistickými znakmi.

Ak chceme vyšetrovať vonkajšiu štruktúru pôvodných premenných, napr. rozdeliťobjekty do skupín, opäť sa úloha zjednoduší, ak sa rieši v priestore menšej dimenzie. Možným riešením sú diskriminačné súradnice.

Ak chceme vyšetriť vnútornú štruktúru premenných,vhodné sú kanonické korelácie.

a) Metóda hlavných komponentov.

Je známe,že metóda je založená na nasledujúcom tvrdení. Ak p-rozmerný náhodný vektor $X\sim(\mu, \Sigma)$, tak existuje ortogonálna matica $B=(b_1,..., b_p)$; $b_j' b_j =1, j=1,..,p$,že $B'\Sigma B = \mathrm{diag}(\lambda_1,...,\lambda_p)$; $\lambda_1 \geq ... \geq \lambda_p \geq 0$; j-ty hlavný komponent vektora X definovaný ako $Y_j = b_j' (X-\mu)$ je ortogonálna projekcia $X-\mu$ v smere vektora b_j.

Metóda hľadá lineárnu transformáciu, ktorá zabezpečí, že niektoré nové premenné majú zanedbateľnú disperziu a teda o súbore objektov neposkytnú žiadnu podstatnú informáciu. Je to zároveň metóda redukcie výberového priestoru: ide o transformáciu vektora X na k-rozmerný vektor $Y_{(k)}$ tak, že pre vhodnú voľbu pxk matice A je $AY_{(k)}$ „ blízko“ k X. Možno predpokladať $\mu=0$,$E(Y_{(k)})=0$ a $D(Y_{(k)})= I_k$. Ako mieru „blízkosti“ použijeme vhodnú funkciu f kovariančnej matice $D(X- AY_{(k)})$. Ide o reálnu funkciu na množine pxp – kladne semidefinitných matíc, ktorá je čiastočne usporiadaná obvyklým spôsobom. Požiadavky na f: striktne rastúca a invariantná vzhľadom na ortogonálne transformácie

(zachovávajú vzdialenosť). Tieto dve vlastnosti majú, napr.st C, det C alebo $(st\ CC')^{\frac{1}{2}}$. Dá sa ukázať, že f má uvedené vlastnosti práve vtedy, ak $f(C)= g(\mu_1(C),...\mu_p(C))$ pre nejakú striktne rastúcu funkciu g každého argumentu, kde μ_i (C) je klesajúca postupnosť vlastných hodnôt matice C.

Ak maticu B rozdelíme na (B_1,B_2) ,$B_1=(b_1,...,b_k)$, tak $f(D(X- AY_{(k)}))$ nadobúda minimum pre $A= B_1$ a teda $AY_{(k)} = B_1 (B_1)'X$.

V praxi μ a Σ odhadujeme z náhodného výberu.Uvedenú teóriu možno použiť na výberovú kovariančnú maticu S. Za všeobecných predpokladov sú všetky jej vlastné hodnoty rôzne. Nevýhodou hlavných komponentov je , že nie sú škálovo invariantné. Ak použijeme miesto S výberovú korelačnú maticu R, jej vlastné hodnoty a vlastné vektory sa od matice S líšia. Potom aj k môže byť iné.

Klúčová je voľba k.V lepšom prípade existuje ostrá deliaca čiara medzi prvými k vlastnými hodnotami a ostatnými.Odporúča sa použiť aj Kaiserovo pravidlo, ale jeho striktné uplatnenie môže spôsobiť problémy.

b) <u>Faktorová analýza</u>

Metóda hlavných komponentov nemá žiadne predpoklady o vektore X. Ak všetky vlastné hodnoty matice Σ sú rôzne, ortogonálna transformácia B je jediná, ak zvolíme znamienko, napr. prvého vlastného vektora. FA predpokladá, že prvky X sa dajú vyjadriť pomocou menšieho počtu neznámych premenných $F_1,...,F_k$ (spoločné faktory)v tvare lineárneho modelu

$$X= \mu + AF +U \qquad (1)$$

kde U sú tzv.špecifické faktory.

Obvyklé prepoklady sú $EF=EU=0$; $cov(F,U) =0, cov(U)=D= diag(d_1,...,d_p)$; od A žiadame, aby v každom stlpci mala aspoň dva nenulové prvky.

Ak kovariančnú maticu spoločných faktorov označíme Φ, máme

$$\Sigma= A\Phi A' + D \qquad (2)$$

Vyjadrenie X v tvare (1) nie je jednoznačné,a preto vždy možno predpokladať, že spoločné faktory sú štandardizované a nekorelované.V takom prípade

$$\Sigma= AA' + D \qquad (3)$$

Máme teda rozhodnúť, či Σ možno vyjadriť v tvare (3), t.j. či existuje diagonálna matica D s kladnými prvkami, že Σ- D je kladne semidefinitná matica hodnosti k. Následne treba odhadnúť A a D.

Problémy FA sú nasledovné: model je nedefinovaný,ak $1/2p(p+1)-p(k+1) <0$.Teda musíme žiadať $1/2p(p+1)-p(k+1) \geq 0$,t.j. $p >2k$. Ani vtedy nie je A jediné: B=AT,kde T je ortogonálna matica, tiež vyhovuje (3). Podobne, ak $h(A)= r<k$,tak $AA' = A_r (A_r)'$, A_r je hodnosti r.V takom prípade model $X=\mu + A_r H+ U$,kde H je vektor r spoločných faktorov, tiež vyhovuje (3).Obvykle sa požaduje ,aby $A'D^{-1}A$ bola diagonálna matica, potom pre dané D je A jediná až na znamienka. Ani v prípade FA nie je jedno, či uvažujeme kovariančnú alebo korelačnú maticu.

Prvky matice A (faktorové náklady) odhadujeme, napr. metódou hlavných faktorov: Σ - D= AA' je kladne semidefinitná s hodnosťou k implikuje existenciu ortogonálnej matice M tak, že $M'(\Sigma - D) M= \Gamma$, čo je diagonálna matica s prvými k kladnými vlastnými hodnotami. Ak M_1 je matica prvých k vlastných vektorov a $\Gamma_1 =diag(\gamma_1 ,...,\gamma_k)$, tak $\Sigma - D= M\Gamma M' = M_1\Gamma_1 M_1' = AA'$,kde $A= M_1 (\Gamma_1)^{\frac{1}{2}}$.Teda potrebujeme poznať odhad D* matice D, vziať prvých k vlastných vektorov matice S- D* a získať odhad $(M_1*) (\Gamma_1*)^{\frac{1}{2}m}$ matice A.

Ako vidieť FA sa od MHK líši tým , že maticu Σ nahradíme maticou Σ- D, resp.korelačnú maticu R maticou R-D (tzv.redukovaná korelačná matica) .

Mierou „ blízkosti“ X a AF môže byť $U´D^{-1}U$. Potom k-rozmerný vektor F, ktorý minimalizuje $U´D^{-1}U$ je

$$F= (A´D^{-1}A)^{-1} A´D^{-1}X \qquad (4)$$

(Bartlettova metóda odhadu faktorového skóre)

c) Kanonické korelácie

Ak rozdelíme náhodný vektor X na dva subvektory $X_{(1)}$ a $X_{(2)}$ s p_1 a $p_2= p- p_1$ zložkami, mierou lineárnej závislosti medzi $X_{(1)}$ a $X_{(2)}$ je maximum z koeficientov korelácie ľubovoľnej lineárnej kombinácie $a´ X_{(1)}$ s ľubovoľnou lineárnou kombinácoiu $b´ X_{(2)}$. Ak $h(\Sigma_{12}) =k$,kde Σ_{12} je kovariančná matica $X_{(1)}$ a $X_{(2)}$, tak k dvojíc kanonických premenných $(U_i ,V_i),i =1,...,k$ má tvar $U_i = a_i´ X_{(1)}$ a $V_i = b_i´ X_{(2)}$, kde a_i sú vlastné vektory $BB´$, b_i sú vlastné vektory $B´B$,$B=(\Sigma_{11})^{-1/2} \Sigma_{12} (\Sigma_{22})^{-1/2}$ zodpovedajúce nenulovým vlastným hodnotám matice $BB´$ alebo $B´B$ (sú rovnaké).

Opäť možno riešiť ako úlohu nájsť k párov lineárnych kombinácií $U=A X_{(1)}$ a $V= B X_{(2)}$, kde A a B majú k lineárne nezávislých riadkov takých, že $A\Sigma_{11}A´= I_k= B\Sigma_{22} B´$, pričom ako mieru „blízkosti“ U a V možno vziať $E\,||U-V||^2$. Výraz nadobúda minimum, ak U a V sú kanonické premenné. Problém možno zovšeobecniť a riešiť pre viac subvektorov náhodného vektora X.

d) Diskriminačné súradnice

Nech n p-rozmerných pozorovaní je rozdelených do m skupín, pričom do i-tej skupiny patrí n_i pozorovaní. Ak X_{ij} znamená j-te pozorovanie v i-tej skupine a $\overline{X_{i.}}$ resp. $\overline{X_{..}}$ aritmetický priemer v i-tej skupine a celkový aritmetický priemer, tak príslušnosť k jednotlivým skupinám možno posúdiť na základe veľkosti matíc

$$W=\sum_{1}^{m}\sum_{1}^{n_i} (X_{ij} - \overline{X_{i.}})(X_{ij} - \overline{X_{i.}})´$$ tzv.matica súčtov štvorcov a súčinov vnútri skupín a

$$B=\sum_{1}^{m}\sum_{1}^{n_i} (\overline{X_{i.}} - \overline{X_{..}})(\overline{X_{i.}} - \overline{X_{..}})´$$ tzv. matica súčtov štvorcov a súčinov medzi skupinami.

Ak λ_j sú vlastné hodnoty matice $W^{-1}B$ so zodpovedajúcimi (vhodne škálovanými) vlastnými vektormi c_j , tak k-rozmerný vektor $Z_{ij} = C´X_{ij}$,kde $C=(c_1,...,c_k)$ má vlastnosť, že A Z_{ij} je pre A= C „najbližšie „ k X_{ij} v zmysle prvej časti tohto príspevku. Prvky vektora Z_{ij} sa nazývajú diskriminačné súradnice.Umožňujú zaradiť objekty do skupín analýzou menšieho počtu k znakov.

Literatúra

Lamoš,F., Potocký,R.: Pravdepodobnosť a matematická štatistika (Štatistické analýzy). Bratislava, Alfa 1989,vyd.UK 1998.

Seber,G.A.F.: Multivariate observations.Wiley,N.York,1984

Doc. RNDr. Rastislav Potocký, CSc., mim. prof.
Katedra aplikovanej matematiky a štatistiky
FMFI UK
Bratislava,Mlynská dolina
potocky@fmph.uniba.sk

Aplikácia klastrovej analýzy pri určovaní bioindikátorov v povodí

Marek Ando

Abstract: Modeling of biological parameters in every case calls for systematic function analysis of the river. Biological parameters of habitat are represented by criteria curves.This article gives information from the first phase of the research, which is aimed to generalize criteria curves. First results from alluvial rivers of Slovakia confirm precondition that it is possible to generalize criteria curves - this helps to simplify the application of criteria curves in another rivers.

Úvod

Najlepším bioindikátorom povodia danej oblasti sú ryby, pretože najlepšie reagujú na rôzne zmeny v prostredí a dobre indikujú zmeny chovania habitatu akvatického prostredia.

Klastrová analýza bola využitá pri zovšeobecňovaní vhodnostných kriviek jednotlivých druhov rýb a následne pomocou ďalších multivariačných štatistických metód sme sa pokúsili definovať vzťahy medzi jednotlivými faktormi charakterizujúcimi habitat.

Zhluková analýza (Cluster analysis) sa zaoberá tým, ako by mali byť objekty zaradené do skupín tak, aby bola čo najväčšia podobnosť v rámci skupín a čo najväčšia rozdielnosť medzi skupinami. (Rimarčík, 2006)

Jedným z hlavných problémov pri zhlukovej analýze je pojem vzájomnej podobnosti objektov a kvantitatávne vyjadrenie tejto podobnosti. Miera podobnosti (Horváth, Gregušová) je nezáporná reálna funkcia *sm*, ktorá každej dvojici vektorov X_q, X_r z E_p priraďuje číslo sm_{qr}, ak pre všetky dvojice X_q, X_r platí:

a) $0 \le sm(X_q, X_r) < 1$ pre $X_q \neq X_r$

b) $sm(X_q, X_r) = 1$

c) $sm(X_q, X_r) = sm(X_r, X_q)$

Poznáme dva základné typy mier podobnosti: koeficienty asociácie a koeficienty korelácie. Ak vychádzame z predstavy, že zhluk je podmnožina množiny objektov, potom vzdialenosť zhlukov môžeme definovať v zmysle vzdialenosti dvoch množín v metrickom priestore. Podľa C. J. Van Rijsbergena (1970) sa dá definícia zhluku napísať takto:
Máme množinu objektov O={O_1, O_2, O_3 …… O_n}, ktorým zodpovedá n vektorov meraní {X_1, X_2 ..... X_n}. Nech *nsm* je miera nepodobnosti objektov. Zhlukom nazývame takú podmnožinu objektov A množiny objektov O, pre ktorú platí

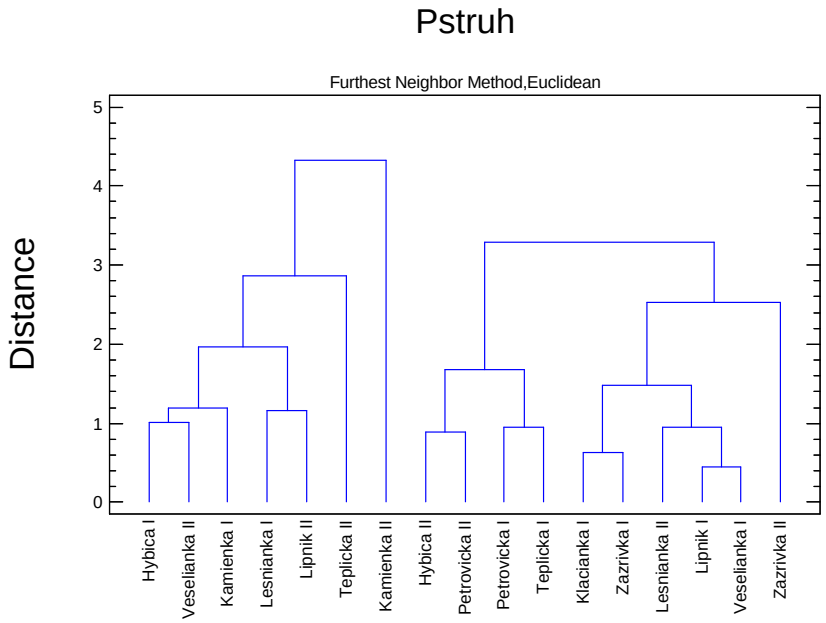
$$\max_{A} \; \mathrm{nsm}(X_q, X_r) < \min_{B} \mathrm{nsm}(X_q, X_s)$$

kde X_q, X_r sú prvky podmnožiny A a podmnožina B je tvorená všetkými dvojicami prvkov, z ktorýcj jeden patrí do podmnožiny A, zatiaľ čo druhý prvok je z doplnku O-A. Teda najväčšia miera nepodobnosti medzi prvkami zhluku musí byť menšia ako miera nepodobnosti medzi každým prvkom zhluku a každým prvkom z množiny objektov, ktorý do zhluku nepatrí. Takáto definícia sa týka konečnej množiny objektov.

Aplikácia

Zhluková analýza bola aplikovaná na pozorovaný štatistický súbor flyšových tokov. Boli vybrané toky podobného charakteru, III. a IV. rádu s veľkosťou povodia 10 až 100 km2. Z pohľadu horninových komplexov bol vybraný flyš. Výber referenčných úsekov bol

realizovaný z databázy 265 povodí, ktoré spĺňali nami definovaný súbor charakteristík. Kvalita habitatu na referenčných úsekov bola indikovaná troma druhmi rýb: pstruh potočný, hlaváč obyčajný a čerebľa pestrá. V mieste odchytu každého jednotlivého kusu daného druhu bola zmeraná hĺbka a rýchlosť prúdenia. Takto bol získaný štatistický súbor, ktorý bol podrobený zhlukovej analýze. Proces bol aplikovaný zvlášť pre každý druh ryby.
Bola zvolená Euklidovská vzdialenosť, použitá bola zhlukovacia metóda Najvzdialenejšieho suseda.
Pre overenie výsledkov zhlukovej analýzy bol súbor rozdelený na dva zhluky, toky upravené a prirodzené. Ako premenné boli použité údaje o hĺbke, rýchlosti a počte jednotlivých druhov rýb.
Pstruh potočný :
Počet kompletných prípadov zhlukovej analýzy bol u pstruha potočného 17.



Na dendrograme môžeme názorne vidieť postup spájania objektov do zhlukov. Objekty, ktoré sa spojili v dolnej časti grafu sú si podobné, tie objekty, ktoré sú spojené vyššie sú rozdielne.

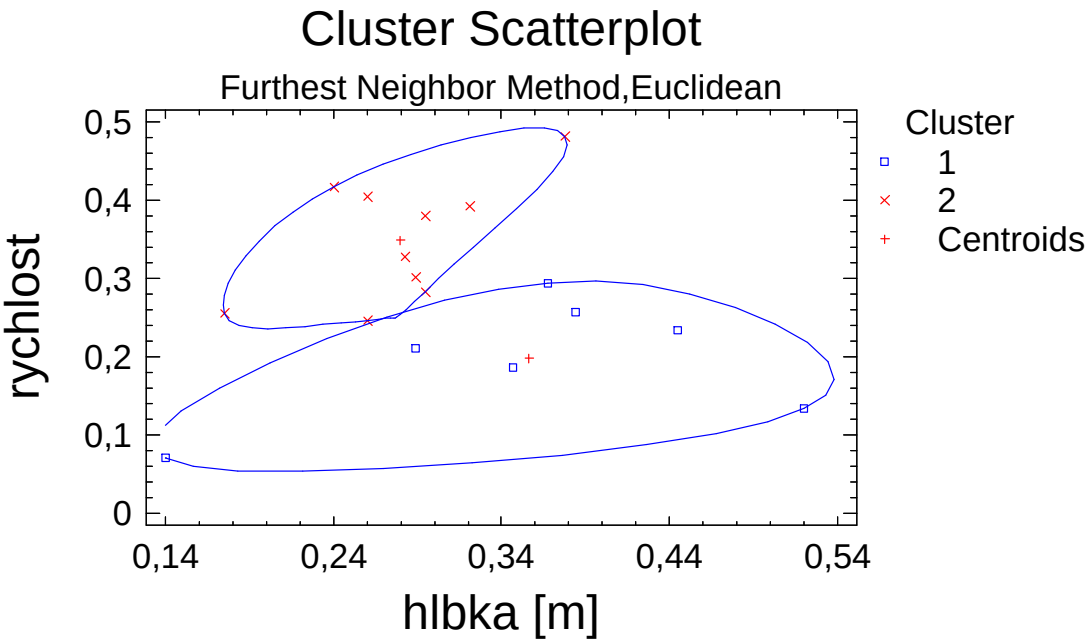
Vytvorili sa dva zhluky:

1) Hybica1, Veselianka2, Kamienka1, Lesnianka1, Lipník2, Tepličká2, Kamienka2

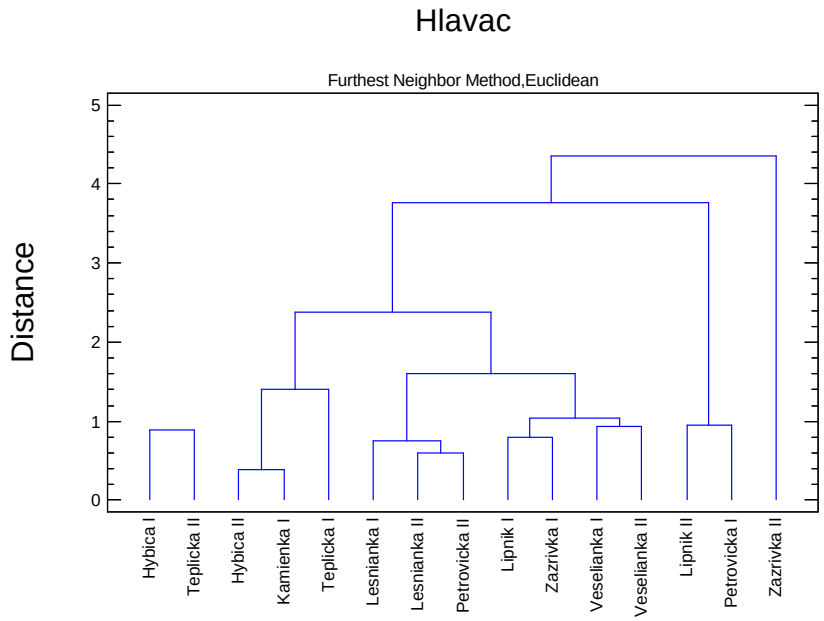
2) Hybica2, Petrovička2, Petrovička1, Tepličká1, Kľačianka1, Zázrivka1, Lesnianka2, Lipník1, Veselianka1, Zázrivka2.

Na zvislej osi je zobrazená vzdialenosť spájania medzi jednotlivými objektami.

Na ďalšom grafe sú charakterizované jednotlivé toky dvoma premennými.



Hlaváč obyčajný:
Počet kompletných prípadov zhlukovej analýzy bol u hlaváča obyčajného 15.



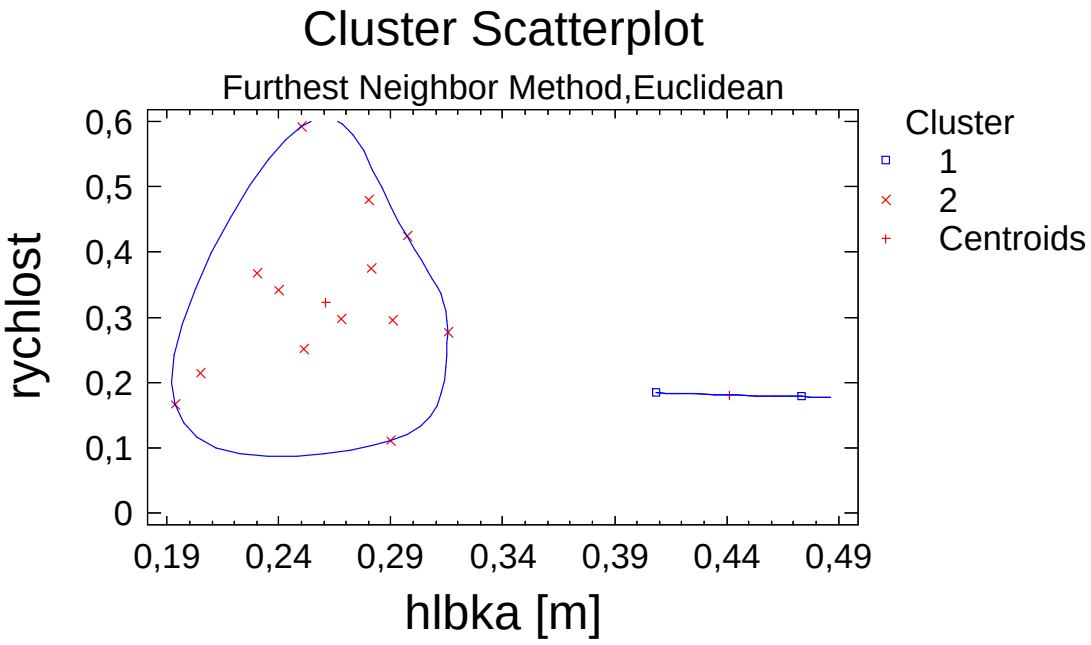
Na dendrograme môžeme názorne vidieť postup spájania objektov do zhlukov. Objekty, ktoré sa spojili v dolnej časti grafu sú si podobné, tie objekty, ktoré sú spojené vyššie sú rozdielne. Vytvorili sa dva zhluky:

1) Hybica1, Tepličká2

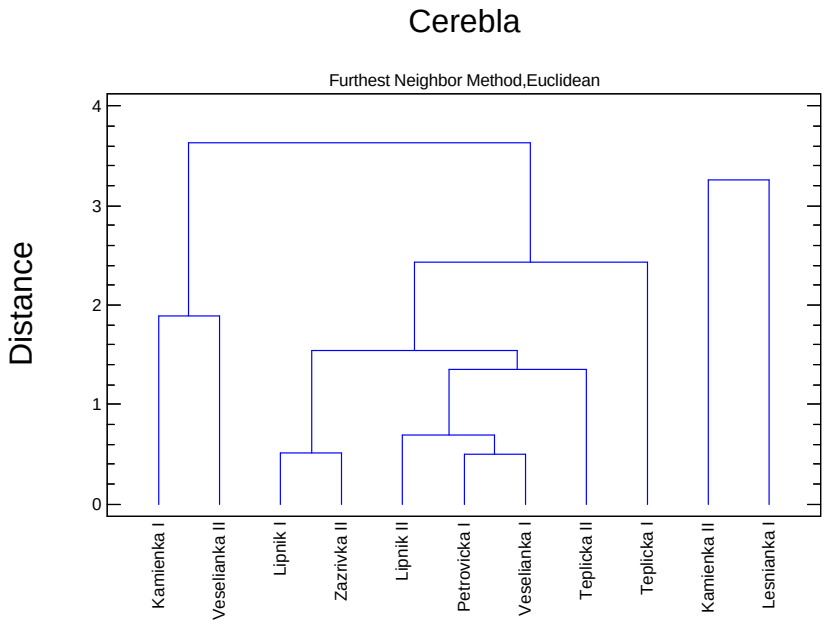
2) Hybica2, Kamienka1, Petrovička2, Petrovička1, Tepličká1, Kľačianka1, Zázrivka1, Lesnianka2, Lipník1, Veselianka1, Zázrivka2., Veselianka1, Veselianka2.

Na zvislej osi je zobrazená vzdialenosť spájania medzi jednotlivými objektami.

Na ďaľšom grafe sú charakterizované jednotlivé toky dvoma premennými



Čerebľa pestrá:
Počet kompletných prípadov zhlukovej analýzy bol u čereble pestrej 11.



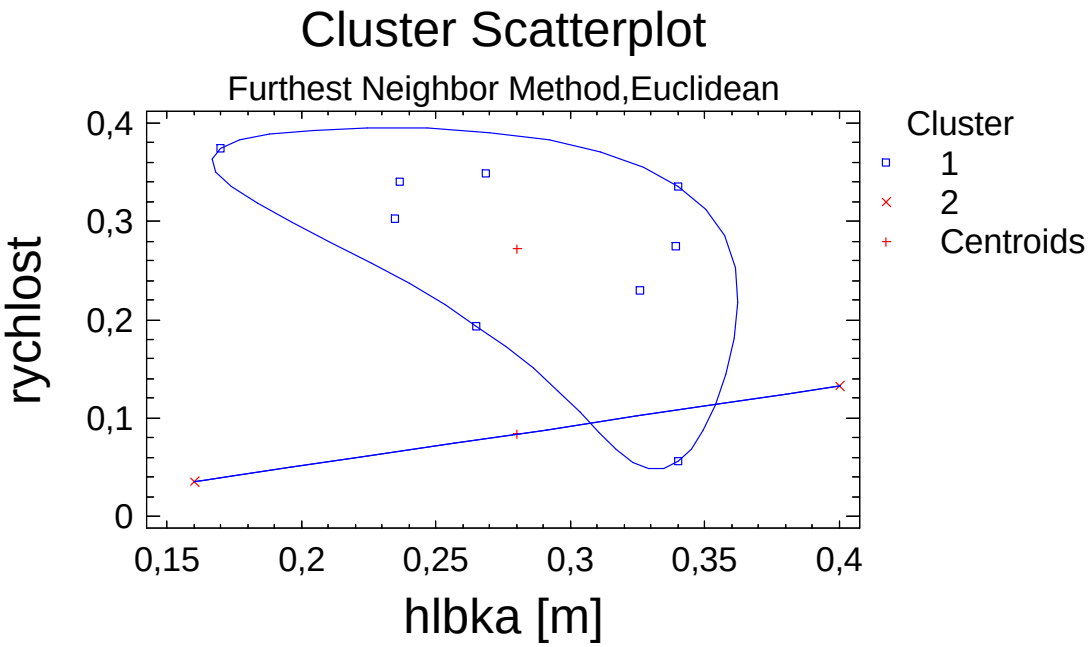
V prípade druhu čereble pestrej sa nám síce vytvorili dva zhluky:

1) Kamienka 2, Lesnianka 1

2) Kamienka 1, Veselianka 2, Lipník 1, Zázrivka 2, Lipník 2, Veselianka 1, Tepličká2, Tepličká 1.

Ako je badateľné z dendrogramu, spájanie objektov do zhlukov prebieha v hornejších polohách grafu, čo indikuje rozdielnosť jednotlivých tokov z pohľadu čereble.

Na nasledujúcom grafe sú charakterizované jednotlivé toky dvoma premennými.



Závery:

Môžeme povedať, že zhluková analýza sa ukazuje ako celkom vhodná metóda pre rozčlenenie jednotlivých tokov a z toho vyplývajúcu charakteristiku pre daný druh ryby. Zovšeobecnenie vhodnostných kriviek jednotlivých druhov rýb a typov tokov umožní výrazné zjednodušenie zberu dát v teréne. Zovšeobecnenie vhodnostných kriviek sa pozitívne prejaví aj na určovaní hydroekologických limitov, vplyvu morfológie koryta na biologické prostredie toku a na optimalizácii tvorby habitatu pri revitalizácii tokov.

Zároveň by bolo vhodné použiť aj iné štatistické metódy z ďalších alternatívnych riešení.

Poďakovanie: Ďakujeme GAV za podporu projektu č. 1/2141/05

Literatúra:

BAKYTOVÁ, H. – UGRON, M. – KONTŠEKOVÁ, O. : Základy štatistiky, 1979. s
HORVÁTH, J., GREGUŠOVÁ, H.: Zhluková analýza a projektívne metódy
van RIJSBERGEN, C., J.: A clustering algorithm. Computer J., 13. 1970. s. 113-115
RIMARČÍK, M.: Základy štatistiky. 2006, 96 s

Mgr. Marek Ando
Katedra vodného hospodárstva krajiny SvF STU
Radlinského 11, 816 68 Bratislava

ando@svf.stuba.sk

Priestorová klasifikácia priemyselných podnikov podľa efektívnosti

Mária Vojtková

Abstract: This article deals with the space classification of industrial enterpises. They are concentrated on charakterization of the economic efficiency. One of procedures that can be used to create classsification is Ward's mininimum variance method.

Úvod

Klasifikácia je základná ľudská pojmová aktivita. Deti sa učia veľmi skoro vo svojom živote klasifikovať objekty svojho okolia a popísať triedy svojim slovníkom. Klasifikácia je tiež fundamentálny proces vedeckej praxe, pretože klasifikačné systémy obsahujú pojmy potrebné pre rozvoj teórie v rámci vedy.

Zhluková analýza zahŕňa širokú paletu metód a postupov, ktoré sa používajú pri riešení problémov typológie objektov a ich klasifikácie. Cieľom zhlukovej anlýzy je rozklad súboru objektov na niekoľko relatívne rovnorodých podmnožín (zhlukov, tried) tak, aby objekty patriace do rôznych zhlukov si boli čo najmenej podobné a objekty patriace do toho istého zhluku si boli podobné čo najviac.

Analýza efektívnosti priemyselných podnikov podľa okresov

Cieľom tohto príspevku je priestorová klasifikácia priemyselných podnikov charakterizovaných vybranými relatívnymi ukazovateľmi. Podkladom sú anonymizované podnikové údaje poskytnuté pre vedecké účely zo Štatisitického úradu SR za rok 2003. Každý podnik je popísaný nasledovnou skupinou ukazovateľov ekonomickej efektívnosti:

- **Z/PH** – rentabilita pridanej hodnoty (ukazovateľ rentability),
- **V/ZAM** – produktivita práce z výnosov (ukazovateľ produktivity),
- **ODP/V** – nákladovosť odpisov (ukazovateľ nákladovosti),
- **ZAV/POH** – koeficient prvej platobnej neschopnosti (ukazovateľ bonity),

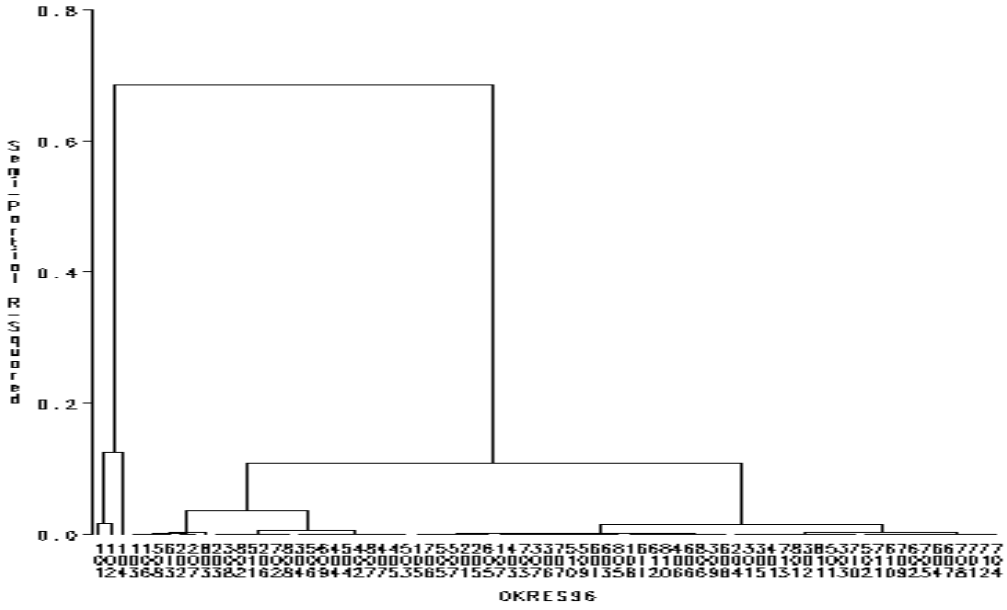
kde Z je výsledok hospodárenia pred zdanením, V sú výnosy, ZAM je priemerný počet zamestnancov v prepočítaných osobách, PH je pridaná hodnota, ZAV sú krátkodobé záväzky a POH sú krátkodobépohľadávky.

Základným predpokladom aplikácie zhlukovej analýzy pri použití Euklidovskej vzdialenosti ako miery nepodobnosti podnikov je predpoklad vzájomnej nezávislosti (nekorelovanosti) ukazovateľov, proti nemu stojí všeobecná súvislosť javov reálneho sveta. Dôsledkom nesplnenia tohto predpokladu je duplicitnosť analyzovaných informácií obsiahnutých vo vstupných dátach, ktorá môže viesť ku značnému skresleniu výsledkov. Splnenie tohto predpokladu bolo zabezpečené dôkladným výberom ukazovateľov efektívnosti (t.j. empirické vylúčenie duplicitných ukazovateľov a ponechanie takých, ktorých vzájomné korelácie sú štatisticky nevýznamné). Iné riešenie je možné dosiahnuť použitím faktorovej analýzy, prípadne komponentnej analýzy pred samotnou klasifikáciou.

Zhlukovacia metóda použitá v tejto analýze je **Wardova metóda**. Pri tejto metóde ide o optimalizačný postup, založený na minimalizácii vnútrozhlukového súčtu štvorcov odchýlok od priemeru zhluku. Spájaním zhlukov do väčších celkov sa tento súčet zvyšuje. Ide o to, aby sa pri spájaní na každom kroku dosiahol minimálny prírastok vnútrozhlukového súčtu štvorcov. Dá sa dokázať, že k minimálnemu prírastku dochádza pri spájaní takých

zhlukov, pre ktoré je štvorec vzdialeností medzi ich priemermi, vážený harmonickým priemerom rozsahov týchto zhlukov, minimálny.

Pre väčšinu s úspechom používaných zhlukových metód je charakteristické to, že ich algoritmy smerujú k realizácii nejakých intuitívnych predstáv o geometrických modeloch hľadaných zhlukov. Výsledkom aplikácie zhlukového algoritmu na danú množinu objektov sú potom zhluky, ktorých grafické zobrazenie pomocou hierarchického stromu obsahuje nasledovný dendrogram.



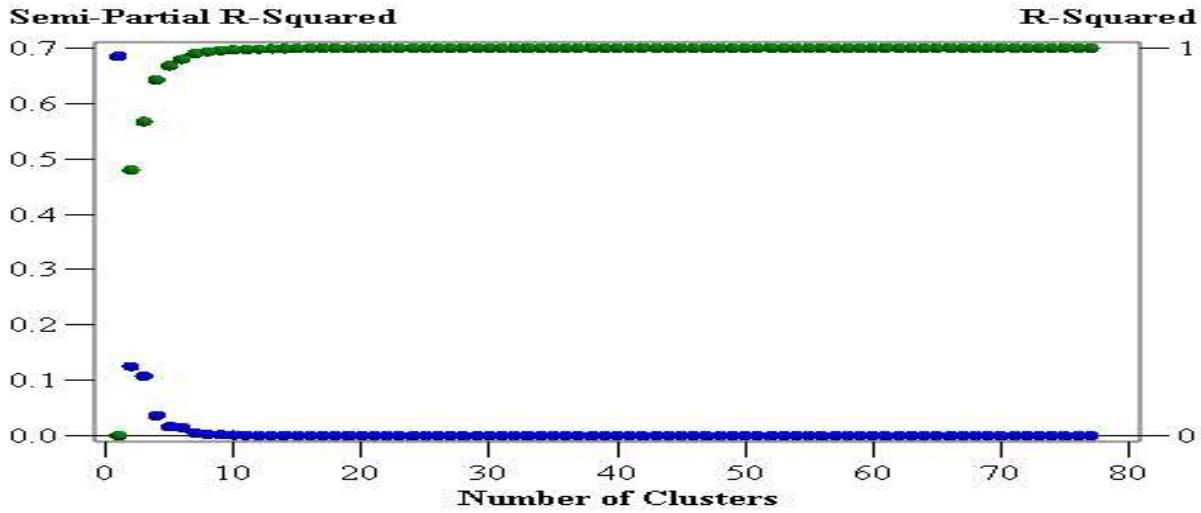
Obr. 1 Dendrogram zhlukovania okresov charakterizovaných ukazovateľmi efektívnosti za jednotlivé priemyselné podniky

Dendrogram je graf znázorňujúci postupnosť zhlukovania okresov. Jeden rozmer reprezentujú čísla okresov a druhý rozmer vzdialenosti medzi nimi. Čiary v poli grafu znázorňujú, ktoré okresy resp. zhluky boli spojené a v akej vzdialenosti. Riešiteľ zhlukovej analýzy sa na základe vzdialenosti rozhodne, kde je bod (zvislá čiara) reprezentujúci situáciu, že vnútroskupinová homogenita je najväčšia a vzdialenosť medzi skupinami najväčšia (medziskupinová homogenita najmenšia). Určenie tohto bodu špecifikuje tiež počet zhlukov.

Na základe predchadzájucej grafickej analýzy a tiež podľa ukazovateľov kvality zhlukovania, ktorých veľkosť v závislosti od počtu zhlukov je znázornená na obr. 2 sme sa ďalej rozhodli súbor rozdeliť do 5 zhlukov. Medzi spomínané charakteristiky patrí:

a) Koeficient determinácie (RSQ) - vyjadruje pomer medziskupinovej sumy štvorcov odchýlok k celkovej sume štvorcov odchýlok. Jeho vyššia hodnota znamená väčšiu medziskupinovú variabilitu, čiže väčšie rozdiely medzi skupinami a tým sú skupiny homogénnejšie. Táto charakteristika sa nachádza v pevnom intervale <0,1> (ak RSQ=0 medzi skupinami nie sú rozdiely, čiže sú rovnaké; ak RSQ=1 ide o maximálne rozdiely medzi skupinami).

b) **Semiparciálny koeficient determinácie** (SPRSQ) – vyjadruje pomer vnútroskupinovej variability (1) mínus vnútorskupinová variabilita (2) k celkovej variabilite. Vnútroskupinová variabilita (1) je vnútroskupinovou variabilitou novovytvoreného zhluku a vnútroskupinová variabilita (2) pozostáva zo súčtu vnútroskupinových variabilít tvoriacich nový zhluk. Tento rozdiel vnútroskupinových variabilít sa tiež označuje ako strata alebo úbytok homogenity. Nízke hodnoty tejto charakteristiky znamenajú, že boli spojené podobné skupiny. Podobne ako predchádzajúca charakteristika i táto sa nachádza v pevnom intervale <0,1>.



Obr.2 Ukazovatele efektivity zhlukovania podľa počtu zhlukov

CLUSTER=1

Obs	OKRES96	OKRES
1	105	Bratislava V
2	107	Pezinok
3	108	Senec
4	201	Dunajská Streda
5	204	Piešťany
6	205	Senica
7	301	Bánovce nad Bebravou
8	302	Ilava
9	303	Myjava
10	305	Partizánske
11	306	Považská Bystrica
12	307	Prievidza
13	309	Trenčín
14	401	Komárno
15	403	Nitra
16	406	Topoľčany
17	501	Bytča
18	502	Čadca
19	505	Liptovský Mikuláš
20	507	Námestovo
21	509	Turčianske Teplice
22	510	Tvrdošín
23	601	Banská Bystrica
24	602	Banská Štiavnica

Obs	OKRES96	OKRES
25	603	Brezno
26	604	Detva
27	605	Krupina
28	606	Lučenec
29	607	Poltár
30	608	Revúca
31	610	Veľký Krtíš
32	611	Zvolen
33	612	Žarnovica
34	701	Bardejov
35	703	Kežmarok
36	704	Levoča
37	705	Medzilaborce
38	706	Poprad
39	707	Prešov
40	708	Sabinov
41	709	Snina
42	710	Stará Ľubovňa
43	711	Stropkov
44	712	Svidník
45	713	Vranov nad Topľou
46	801	Gelnica
47	805	Košice IV
48	806	Košice-okolie
49	810	Spišská Nová Ves
50	811	Trebišov

CLUSTER=2

Obs	OKRES96	OKRES
51	203	Hlohovec
52	206	Skalica
53	304	Nové Mesto nad Váhom
54	308	Púchov
55	402	Levice
56	404	Nové Zámky
57	405	Šaľa
58	407	Zlaté Moravce
59	503	Dolný Kubín
60	504	Kysucké Nové Mesto
61	506	Martin
62	511	Žilina
63	609	Rimavská Sobota
64	702	Humenné
65	802	Košice I
66	807	Michalovce
67	808	Rožňava

CLUSTER=3

Obs	OKRES96	OKRES
68	103	Bratislava III
69	106	Malacky
70	202	Galanta
71	207	Trnava
72	508	Ružomberok
73	613	Žiar nad Hronom
74	803	Košice II

CLUSTER=4

Obs	OKRES96	OKRES
75	101	Bratislava I
76	102	Bratislava II

CLUSTER=5

Obs	OKRES96	OKRES
77	104	Bratislava IV

Obr. 3 Priestorová klasifikácia priemyselných podnikov podľa zhlukov

Cieľom bolo zoskupiť okresy charakterizované priemyselnými podnikmi podľa ukazovateľov efektívnosti do zhlukov tak, aby okresy patriace do toho istého zhluku si boli v istom zmysle podobné.

Záver

Výsledné hodnotenie jednotlivých zhlukov je možné uskutočniť na základe veľkosti ich zhlukových priemerov podľa tabuľky 1.

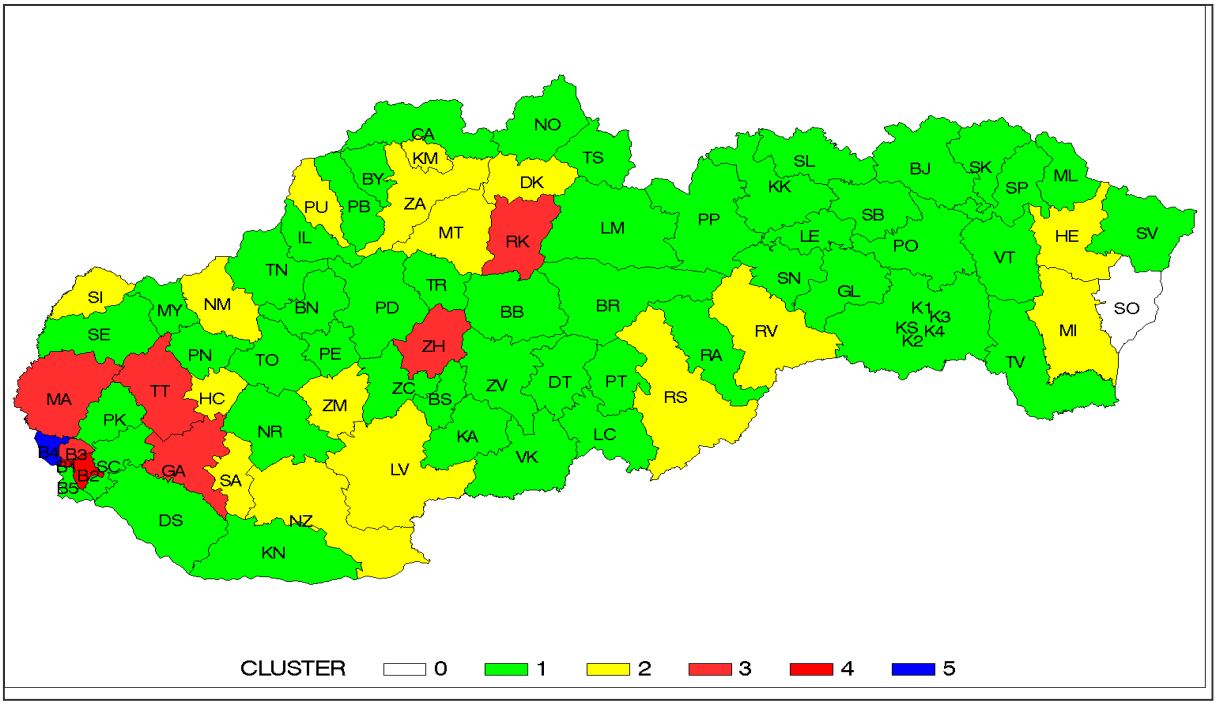
Tab. 1 Zhlukové centroidy ukazovateľov efektívnosti priemyselných podnikov podľa okresov

Zhluky	Početnosť zhlukov	Ukazovatele efektívnosti			
		ZAV/POH	ODP/V	Z/PH	V/ZAM
1	50	1,329	0,046	-0,073	1289
2	17	1,012	0,045	0,142	2473
3	7	0,888	0,051	0,126	3965
4	2	1,293	0,046	0,419	7638
5	1	0,900	0,030	0,344	16688

1. zhluk – zhluk, ktorý zahŕňa najväčší počet okresov je charakteristický priemyselnými podnikmi s vysokou prvotnou platobnou neschopnosťou, druhou najvyššou nákladovosťou odpisov a zápornou mierou podielu vlastníka na rozdelení vlastných výsledkoch podniku. Tento zhluk dosahuje aj najnižšiu hodnotu hrubého

ekonomického prínosu na jedného zamestnanca. Celkovo tento zhluk zahŕňa okresy s priemyselnými podnikmi s najnižšou ekonomickou efektívnosťou.

2. zhluk – v okresoch patriacich do tohoto zhluku sú priemyselné podniky so strednou hodnotou platobnej neschopnosti, nákladovosti odpisov a rentability pridanej hodnoty. Priemerný ukazovateľ produktivity práce z výnosov dosahuje druhú najnižšiu hodnotu.
3. zhluk – zahŕňa okresy s priemyselnými podnikmi s najnižšou prvotnou platobnou neschopnosťou na jednej strane, ale s najvyššou nákladovosťou odpisov na strane druhej. Priemerný ukazovateľ produktivity práce je na strednej priečke a rentabilita pridanej hodnoty na druhej najnižšej.
4. zhluk – obsahuje dva Bratislavské okresy, v ktorých sa nachádzajú priemyselné podniky s druhou najvyššou prvotnou platobnou neschopnosťou, nákladovosťou odpisov, ale súčasne s najvyššou mierou podielu vlastníka na rozdelení vlastných výsledkoch a druhým najvyšším ekonomickým prínosom na zamestnanca.
5. zhluk – ide o samostatný Bratislavský okres, ktorý charakterizujú priemyselné podniky s druhou najnižšou prvotnou platobnou neschopnosťou a najnižšou nákladovosťou výnosov. Priemerná produktivita práce z výnosov tu dosahuje najvyššiu hodnotu a priemerná rentabilita pridanej hodnoty druhú najvyššiu hodnotu, čiže celkovo môžeme tento zhluk zhodnotiť, ako zhluk s priemyselnými podnikmi s najvyššou ekonomickou efektívnosťou.



Obr. 4 Mapa okresov podľa zhlukov charakterizovaných ekonomickou efektívnosťou priemyselných podnikov

Grafické znázornenie priestorového rozmiestnenia priemyselných podnikov charakterizovaných vybranými ukazovateľmi efektívnosti obsahuje hore uvedená mapa. Zhluk 0, ktorý je znázornený na obrázku zahŕňa okresy, ktoré nevykazujú žiadne priemyselné podniky nad 20 zamestnancov, avšak sú súčasťou mapy SR.

Literatúra

[1] ALDENDERFER, M. S. - BLASHFIELD, R. K.: *Cluster analysis*. Beverly Hills: SAGE, 1985.
[2] ANIL, J. K. – DUBES, R.: *Algorithmus for clustering data*. Englewood Cliffs, New Jersey: Prentice hall, 1988.
[3] CHAJDIAK, J.: *Základy hospodárskej štatistiky*. Bratislava: Statis, 2004.
[4] KHATTREE, R. – NAIK, N. D.*: Applied Multivariate Statistics with SAS® Software.* First edition, Cary, NC: SAS Institute Inc., 1995.
[5] SAS Institute Inc.: SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 1, Cary, NC: SAS Institute Inc., 1989.
[6] SHARMA, S.: *Applied multivariate techniques*. New York: John Wiley & Sons, 1996.
[7] ZALAI, K. a kol.: *Finančno-ekonomická analýza podniku*. Bratislava: Sprint, 1998.

Kontakt:
Ing. Mária Vojtková, PhD.
Katedra štatistiky, FHI, EU Bratislava
E-mail: vojtkova@euba.sk

Genetický polymorfizmus a markery poškodenia DNA.

Ladislava Wsólová, Zuzana Džupinková, Mária Dušinská

Abstract

Aging is an inevitable consequence of being a multicellular organism, associated with a decline in physiological function and resulting in death. Aim of our project "THE ROLE OF GENETIC PREDISPOSITION, IMMUNE SYSTEM AND NUTRITION IN AGING" is to show if and how the polymorphisms in DNA repair genes can influence repair processes in human and if they have any relation to human aging.

We collected 291 healthy volunteers: 151 young persons (18 - 25 years) and 140 seniors (63 - 70 years). The young group consisted of 73 men and 78 women, the old group consisted of 39 men and 101 women.

One of the measured types of DNA damage, altered purines (FPG-sensitive sites) in lymphocyte DNA was higher in young group compared with seniors, especially in young women. Together with polymorphisms in some DNA repair genes, the highest level of FPG-sites was observed in the subjects with wild type of *NBS1* gene and simultaneously with wild type of *XPC* gene in intron 9.

Starnutie ľudskej populácie patrí v súčasnosti medzi často diskutované témy. Na svete žije oproti minulosti oveľa viac starých ľudí. Tento trend pozorujeme od II. polovice 20. storočia a má i naďalej stúpajúci charakter (Kouřilová J., 2003). Náš život sa stal v priebehu stáročí bezpečnejším vďaka mnohým zdravotníckym opatreniam, ale napriek tomu nie sme schopní zaistiť, aby naše telesné schránky nechátrali.

Existujú značné rozdiely v názoroch na to, prečo organizmy vôbec starnú a akékoľvek definovanie starnutia je veľmi náročné (Helfand L. S., a Rogina B., 2003). Preto bolo navrhnuté veľké množstvo teórií o príčinách starnutia. Jednu skupinu tvoria vývojovo –genetické teórie, ale v súčasnosti je zrejmé, že pôsobenie génov a prostredia nemožno hodnotiť oddelene. Môžeme tvrdiť, že na starnutí sa podieľajú všetky gény jedinca - genofond získaný od rodičov i aktuálny stav genómu, ktorý odráža chyby pri replikácii DNA spolu s vplyvmi životného štýlu a environmentálnych faktorov.

Odpoveď na otázku „prečo starneme?" sme sa pokúšali nájsť aj v rámci projektu „ÚLOHA GENETICKEJ PREDISPOZÍCIE, IMUNITNÉHO SYSTÉMU A VÝŽIVY V PROCESE STARNUTIA". Na sledovanie procesov starnutia sme použili multidisciplinárny a molekulárno-epidemiologický prístup. Analyzovali sme úlohu výživy, životného štýlu, imunitnej odpovede organizmu a individuálnej predispozície jedinca (genetické polymorfizmy vo vybraných génoch) na úrovni bunkových markerov a ich vzťah k procesu starnutia. Našou úlohou bolo získať informácie o bunkových a molekulárnych základoch starnutia pomocou sledovania vhodných biomarkerov u dvoch vekovo rozdielnych vyvážených skupín obyvateľstva. Medzi genetické faktory, ktoré môžu modulovať starnutie patrí i genetická variabilita (polymorfizmus) identifikovaná v mnohých génoch

kódujúcich systémy priamo či nepriamo zabezpečujúce stabilitu a integritu DNA. Sú to gény kódujúce proteíny zapojené do opravy DNA a gény pre enzýmy biotransformácie exogénnych látok (xenobiotík). Projekt bol zameraný na identifikáciu možných vzťahov medzi molekulárnymi markermi, psychologickými a fyzickými aspektmi starnutia, ako aj na identifikáciu rozdielov v starnutí medzi pohlaviami.

Vieme, že genetické faktory, ktoré podmieňujú vnímavosť jedinca k určitým ochoreniam, vykazujú u rôznych populácií rozdielnu frekvenciu výskytu. K faktorom životného štýlu a prostredia sa zaraďujú riziká, ktoré zvyšujú možnosť výskytu určitých ochorení, ako napr. konzumácia alkoholu, fajčenie, choroby z povolania, preľudnenie, rozdielna ekonomická úroveň a výživa.

Jedna z teórií považuje za hlavnú príčinu starnutia akumuláciu oxidačného poškodenia v biomolekulách (DNA, proteíny a lipidy) (Harman D., 1956), čo môže byť zapríčinené i nedostatočným príjmom antioxidantov vo výžive. Hromadenie oxidačného poškodenia v lipidoch a proteínoch ohrozuje najmä bunky v tkanivách s dlhou životnosťou. Oxidačné poškodenie DNA môže mať závažnejšie následky. Jedným z nich je trvalá zmena DNA - mutácia, ktorá môže spôsobiť zánik alebo v horšom prípade modifikáciu vlastností bunky (napr. vznik nádorov). Poškodenie DNA je výsledkom pôsobenia poškodzujúcich agens na jednej strane, rýchlosti a účinnosti obranných procesov na strane druhej. Rozsah poškodenia limituje najmä presnosť opravy poškodenej DNA ako i ďalšie obranné enzymatické mechanizmy (napr. enzýmy kataláza, superoxiddismutáza, glutatiónperoxidáza, tripeptid glutatión) spolu s antioxidantmi prijímanými v potrave, ako sú napr. vitamín C, vitamín E, karotenoidy a mnohé ďalšie známe i neznáme molekuly.

Poškodenie DNA (jednovláknové zlomy, modifikované puríny a pyrimidíny) bolo merané pomocou metódy comet assay na vitálnych lymfocytoch izolovaných z venóznej krvi Parameter FPG- sensitive sites (oxidované puríny) odráža aktuálny stav poškodenia DNA v bunke a môže sa v čase pomerne dynamicky meniť v závislosti od exo- i endogénnych vplyvov a reparačného systému bunky. Proteín kódovaný *XPC* génom sa spolu s ďalším proteínom podieľa na rozpoznávaní poškodenia a tým je zodpovedný za spustenie jednej z opravných dráh DNA – nukleotidová excízna oprava (Sugasawa K. a kol., 1998). Proteín kódovaný génom *NBS1* sa uplatňuje v prvých fázach opravy dvojvláknových zlomov DNA pomocou opravnej dráhe nazývanej homologická oprava (Trujillo K. M. a kol., 1998). Genetická variabilita vo vybraných génoch bola stanovovaná pomocou metódy PCR-RFLP.

Zisťovali sme, s ktorými premennými môže byť asociované poškodenie DNA. Na postavenie hypotéz sme použili modul AnswerTree. Ďalej sme pracovali v SPSS 13. 0. Vzťah medzi poškodením DNA a niektorými premennými (veková skupina, pohlavie, polymorfizmus v génoch *XPC* a *NBS1*) je charakterizovaný nasledovným matematickým modelom:

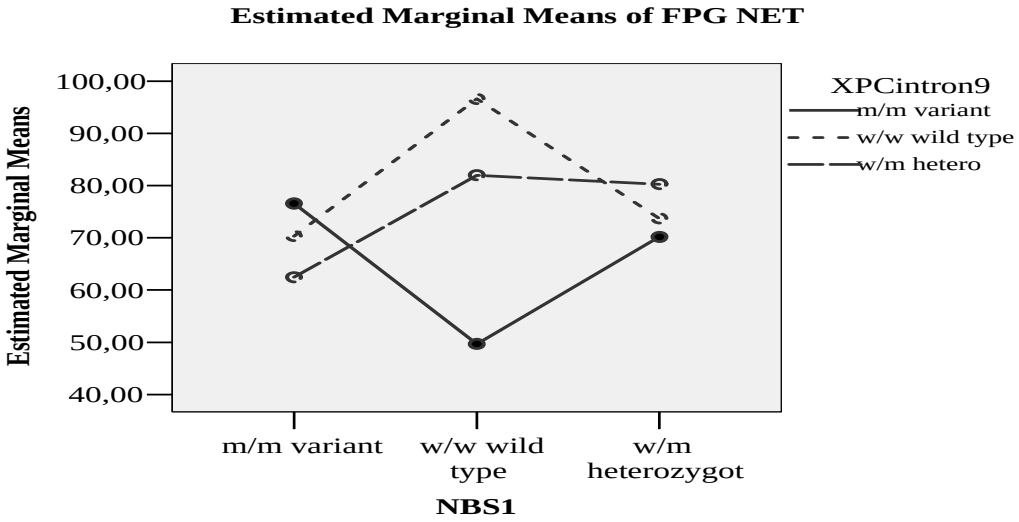
Tests of Between-Subjects Effects

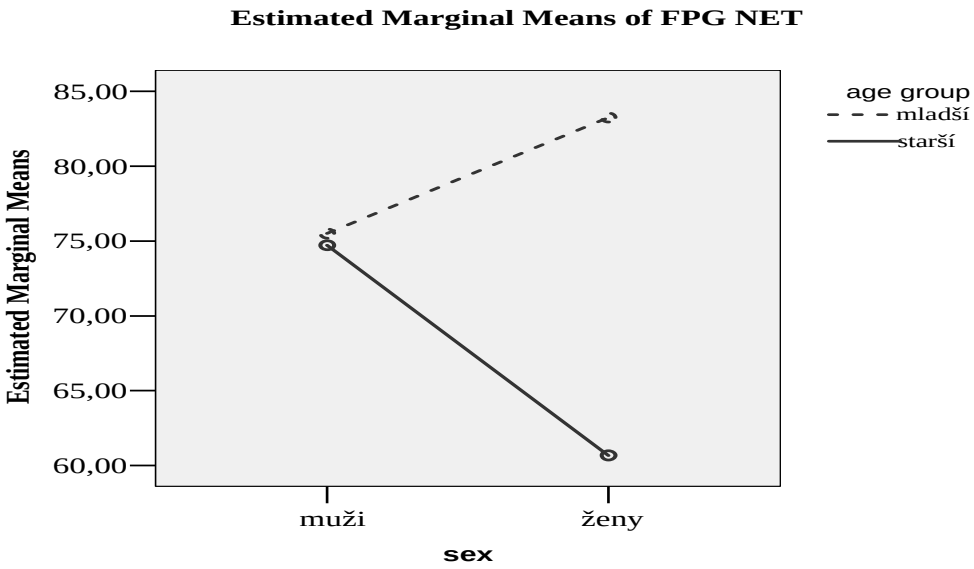
Dependent Variable: FPG NET

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	70956,505[a]	11	6450,591	3,488	,000
Intercept	637941,68	1	637941,683	344,909	,000
XPCintron9 * NBS1	47461,222	8	5932,653	3,208	,002
age group * sex	23044,441	3	7681,480	4,153	,007
Error	508638,80	275	1849,596		
Total	2295841,7	287			
Corrected Total	579595,31	286			

a. R Squared = ,122 (Adjusted R Squared = ,087)

Závislá premenná **FPG NET** vyjadruje aktuálne poškodenie DNA. Nezávislá premenná **age group** znamená príslušnosť k vekovej skupine (0=mladší, 1=starší), **sex** vyjadruje pohlavie (0=muži, 1=ženy), ***XPC*intron9** a ***NBS1*** sú označenia génov, v ktorých boli polymorfizmy sledované. Oba majú 3 formy genotypov: m/m (mutantný typ), w/m (heterozygót) a w/w (divý typ). Model ukazuje, že poškodenie DNA je štatisticky významne ovplyvňované interakciou medzi jednotlivými genotypmi a interakciou medzi vekovou skupinou a pohlavím.





Oxidačné poškodenie DNA, ktorého markery i my sledujeme, je preferenčne opravované bázovou excíznou opravou. Interakcia medzi nami sledovanými markermi oxidačného poškodenia DNA a polymorfizmami v génoch zapojených do iných opravných dráh však nie je až tak prekvapujúca. Podobná interakcia sa ukázala i pri štatistickom vyhodnotení inej štúdie (Dušinská M. a kol., 2006). Nukleotidová excízna oprava, ktorá hlavne opravuje rozsiahlejšie poškodenia DNA, sa za určitých podmienok, môže byť zapojená i do opravy poškodených báz. Interakcia medzi génmi z dvoch odlišných opravných dráh a jej vplyv na marker poškodenia naznačuje skutočnosť, že na oprave určitého typu poškodenia DNA sa môže podieľať viacero mechanizmov, ktoré vzájomne interagujú.

Prirodzene by sme očakávali, že divý variant génov bude pre jedinca benefitom. Mnohé populačné štúdie zamerané na výskum polymorfizmov však ukazujú, že nie vždy to, čo je považované za divý typ, musí predstavovať funkčne výhodný variant, a to najmä v kombinácii s rôznymi variantmi ďalších génov.

Model ďalej ukazuje, že seniori majú priemerne nižšie aktuálne poškodenie DNA ako mladší jedinci, štatisticky signifikantne sa to prejavilo v skupine žien. Všeobecne prijímaná hypotéza predpokladá pokles opravnej kapacity DNA a následný nárast DNA poškodenia s vekom. Nami zistený pokles poškodenia DNA u seniorov by sa dal vysvetliť zmenou životného štýlu u starších ľudí, ktorý môže mať vo vyššom veku závažnejší vplyv než genetická výbava jedinca. Prirodzený výber môže pôsobiť len v reprodukčnom období jedinca, keď sa rozhoduje o ďalšom prežití druhu, a preto vplyv génov exprimovaných v neskorších vývinových štádiách jedinca nezohráva veľmi veľkú úlohu (Hamet P., Tremblay J., 2003).

Záver: Výsledný model naznačuje, že poškodenie DNA (FPG-sensitive sites) je ovplyvňované kombináciou jednotlivých foriem génu *XPC* v intróne 9 a génu *NBS1*. Najvyššie hodnoty poškodenia sú pri divom type (w/w) oboch sledovaných polymorfizmov. Najnižšie poškodenie bolo zistené pri kombinácii divého typu (w/w) génu *NBS1* a mutatntného typu (m/m) intrónu 9 v géne *XPC*. Poškodenie ďalej závisí od kombinácie medzi vekovou skupinou a pohlavím,

najvyššie hodnoty poškodenia boli zistené u skupiny mladých žien. Skutočnosť, že tento marker oxidačného poškodenia DNA sa viaže s dvomi interakciami medzi odlišnými premennými je pravdepodobne daný rozdielnou biologickou podstatou daných premenných a ich závažnosťou, s akou vplývajú na vybraný biomarker. I tieto faktory treba brať do úvahy pri interpretácii experimentálnych biologických výsledkov.

Projekt bol podporený grantom APVT-21-013202.

Použitá literatúra:

DUŠINSKÁ, M., DŽUPINKOVÁ, Z., WSÓLOVÁ, L., HARRINGTON, V., COLLINS, A. R. (2006). Possible involvement of XPA in repair of oxidative DNA damage deduced from analysis of damage, repair and genotype in a human population study. *Mutagenesis, in press.*

HAMET, P., TREMBLAY, J. (2003). Genes of Aging. *Metabolism 52* (*10 Suppl 2*): 5 – 9.

HARMAN, D. (1956). Aging: a theory based on free radical and radiation chemistry. *J Gerontol 11*: 298 – 300.

HELFAND, S. L., ROGINA, B. (2003). Molecular genetics of ageing in the fly: is this the end of the beginning? *BioEssay 25*: 134-141.

KOUŘILOVÁ, J. (2003). Zdravé starnutie.
http://www.szulm.sk/miscel/starnutie.html. 23. 3. 2004

STEVEN, N. A. (1997). Why we age. John Wiley & Sons, Inc., 256 s.

SUGASAWA, K., NG, J. M., MASUTAMI, C., IWAI, S., VAN DER SPEK, P. J., EKER, A. P., HANAOKA, F., BOOTSMA, D., HOEIJMAKERS, J. H. (1998). Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol. Cell*: 223 – 232.

TRUJILLO, K. M., YUAN, S.-S. F., LEE, E. Y.-H. P., SUNG, P. (1998). Nuclease activities in a complex of human recombination and DNA repair factors Rad50, Mre11, and p95. *J. Biol. Chem. 273*: 21447 – 21450.

Adresa: RNDr. Ladislava Wsólová, Slovenská zdravotnícka univerzita, Limbová 12, 83303 Bratislava, ladislava.wsolova@szu.sk

Význam analýzy hlavných komponentov pri riešení úloh viacrozmernej štatistiky

Silvia Megyesiová

Abstract:

The article deals with the problem of a strong linear correlation between some selected and analyzed indicators. A strong linear correlation is not acceptable for example by doing some multivariate statistics analysis or by creating econometrics models. That is why we decided to explain the process of main components analysis by a simply example of data, that are very often used by analysis of economic situation in different countries.

Úvod

V niektorých prípadoch sa stretávame so situáciou, keď východzí počet sledovaných štatistických znakov na skúmaných štatistických jednotkách, objektoch je veľký a pre interpretáciu neprehľadný resp. niektoré štatistické znaky sú vzájomne lineárne závislé, silne korelované. Pre zjednodušenie analýz ako i interpretácií výsledkov je preto vhodné preskúmať možnosť, či by štatistické znaky nebolo možné nahradiť menším počtom znakov a to tak, aby pritom nedošlo k veľkej strate informácií.

Na riešenie problému zníženia dimenzie pôvodnej matice pozorovaní ako i odstránenie silnej vzájomnej korelácie medzi pôvodnými premennými slúžia dve formálne podobné metódy viacrozmernej štatistickej analýzy: *komponentná analýza* a *faktorová analýza*. Obe tieto metódy vychádzajú z pôvodného súboru nameraných premenných a na základe analýzy ich kovariančných resp. korelačných matíc ich tranformujú do menšieho počtu nových, latentných premenných.

Podstatou metódy hlavných komponentov je požiadavka, aby nové premenné vyčerpali maximum celkového rozptylu pôvodných premenných resp. aby maximálne reprodukovali celkovú kovariačnú maticu pôvodných premenných.

Úlohou faktorovej analýzy je požiadavka, aby nové premenné maximálne reprodukovali nediagonálne prvky kovariačnej matice pôvodných premenných, teda aby vysvetľovali závislosti medzi premennými.

Cieľom analýzy hlavných komponentov je teda redukcia dimenzie riešeného problému. Od nových, latentných premenných sa pritom požaduje, aby dostatočne presne reprezentovali

pôvodne napozorované premenné a aby nové premenné vyčerpali maximum celkového rozptylu pôvodných premenných.

Riešením úlohy hlavných komponentov sa dostávame k modelovému vyjadreniu pôvodných, priamo merateľných premenných menším počtom hypotetických skrytých komponentov, ktoré nie je možné priamo merať.

Vzťahy medzi pôvodnými napozorovanými veličinami a komponentami sa vyjadrujú sústavou lineárnych rovníc. Parametre týchto rovníc, tzv. komponentné váhy nám umožňujú interpretáciu významu jednotlivých komponentov.

Hlavné komponenty sú zoradené podľa dôležitosti, teda podľa klesajúceho rozptylu (od najväčšieho po najmenší). Väčšina informácií o variabilite pôvodných údajov je pritom sústredených do prvej komponenty a najmenej informácií je obsiahnutých v poslednej komponente.

Potreba nahradenia pôvodných premenných niekoľkými hlavnými komponentami je často uplatňovaná pre riešení ekonometrických úloh, keď počet premenných prevyšuje počet pozorovaní ako i pri riešení úloh, ktoré vyžadujú nekorelovanosť premenných, pretože nové, latentné hlavné komponenty sú vzájomne nekorelované.

Z údajov matice pozorovanív **X** typu *n x p* určíme kovariačnú (**S**) resp. korelačnú maticu (**R**) v prípade, že pri výpočte hlavných komponentov vychádzame z veličín napozorovaných v rovnakých merných jednotkách, v opačnom prípade je potrebné pred samotným výpočtom kovariačnej resp. korelačnej matice prepočítať namerané hodnoty matice X na normovaný tvar.

1. Korelačná analýza vybraných ukazovateľov

Hlavné komponenty určené z kovariačnej matice **S** a z korelačnej matice **R** sú rôzne, väčšina aplikácií analýzy hlavných komponentov vychádza z použitia korelačnej matice. Ak sú premenné vyjadrené v rovnakých merných jednotkách vhodnejšie je použitie kovariačnej matice.

Uveďme si výber hlavných komponentov na konkrétnom príklade analýzy deviatich ukazovateľov, ktoré boli zistené za všetky krajiny EÚ za rok 2003:

A1 – hrubý domáci produkt na obyvateľa v parite kúpnej sily,

A2 – koeficient rastu hrubého domáceho produktu v stálych cenách,

A3 – harmonizovaný index spotrebiteľských cien,

A4 – deflátor hrubého domáceho produktu,

A5 – dlhodobá miera nezamestnanosti v %,

A6 – koeficient rastu celkového počtu zamestnaných,

A7 – tvorba hrubého fixného kapitálu (% z HDP),

A8 – podiel vývozu a dovozu v %,

A9 – verejný dlh (% z HDP).

Na základe korelačnej analýzy zistíme silu lineárnej závislostí medzi jednotlivými sledovanými kritériami teda štatistickými znakmi.

Jednotlivé párové koeficienty korelácie charakterizujú nízku až stredne silnú koreláciu medzi jednotlivých ukazovateľmi.

Najvyššia hodnota párového koeficienta korelácie je medzi ukazovateľmi A_2-A_6 a to na úrovni 0,636. Na základe tohto koeficienta korelácie medzi ukazovateľmi A_2 a A_6 môžeme skonštatovať, že existuje lineárna priama stredne silná závislosť medzi koeficientom rastu HDP a koeficientom rastu celkového počtu zamestnaných.

Medzi ukazovateľmi, ktoré charakterizujú cenový vývoj v jednotlivých krajín EÚ, konkrétne medzi ukazovateľom A3 a A4 bol dosiahnutý párový koeficient korelácie na úrovni 0,379. Tento nízky koeficient korelácie medzi harmonizovaným indexom spotrebiteľských cien a deflátorom hrubého domáceho produktu poukazuje na správnosť zaradenia oboch indexov do našej analýzy, pretože ich koeficient korelácie je veľmi nízky.

Tab. 1 Párové koeficienty korelácie jednotlivých kritérií

Pearson Correlation Coefficients, N = 25									
	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**
A1	1.00000	-0.43453	-0.12357	0.05745	-0.59972	-0.05611	-0.42548	0.59115	0.04743
A2	-0.43453	1.00000	-0.01608	-0.20866	0.40753	0.63605	0.47575	-0.47392	-0.52541
A3	-0.12357	-0.01608	1.00000	0.37894	0.20785	0.21436	0.29752	-0.08625	0.03310
A4	0.05745	-0.20866	0.37894	1.00000	-0.27700	0.33202	0.12895	-0.28823	0.25013
A5	-0.59972	0.40753	0.20785	-0.27700	1.00000	0.09405	0.30384	-0.27566	0.01207
A6	-0.05611	0.63605	0.21436	0.33202	0.09405	1.00000	0.55149	-0.47404	-0.24844
A7	-0.42548	0.47575	0.29752	0.12895	0.30384	0.55149	1.00000	-0.62879	-0.27528
A8	0.59115	-0.47392	-0.08625	-0.28823	-0.27566	-0.47404	-0.62879	1.00000	0.04312
A9	0.04743	-0.52541	0.03310	0.25013	0.01207	-0.24844	-0.27528	0.04312	1.00000

V prípade, že nami analyzované štatistické znaky sú navzájom silne korelované je ich potrebné v ďalších analýzach nahradiť nekorelovanými, latentnými premennými tak, aby

pritom nedošlo k veľkej strate informácií. Tieto nové nemerateľné premenné získame pomocou analýzy hlavných komponentov.

2. Praktická ukážka výberu hlavných komponentov

K analýze hlavných komponentov opäť použijeme štatistický softvér, ktorý nám poskytne charakteristiky umožňujúce rozhodnutie o počte potrebných hlavných komponentov, ktoré nahradia naše pôvodné štatistické znaky a zároveň nám vypočítajú tzv. komponentné skóre, teda hodnoty jednotlivých hlavných komponentov.

Analýza hlavných komponentov bola uskutočnená pomocou korelačnej matice normovaných hodnôt premenných uvedených v matici **X** bez rotácie faktorov.

V tabuľke 2 sú uvedené charakteristické čísla korelačnej matice *Eigenvalues* v stĺpci s označením *Total* . Súčet týchto vlastných čísel je rovný počtu analyzovaných štatistických znakov, teda v našom prípade 9.

Ďalšie dva stĺpce uvádzajú podiel vysvetlenej variability daným komponentom na celkovej variabilite v percentách pomocou relatívnej početnosti a kumulatívnej relatívnej početnosti.

Vychádzajúc z tabuľky 32 je možné rozhodnúť pomocou vzťahov *3.58-3.60* ako i odporúčaní v literatúre o potrebnom počte hlavných komponentov, ktoré by mali nahradiť naše pôvodné premenné:

- z údajov v stĺpci *Total* vyberáme také hlavné komponenty, ktorých vlastné číslo korelačnej matice je vyššie než 0,7. Na základe tohto pravidla by sme ako potrebný počet hlavných komponentov určili 4 hlavné komponenty,

Tab. 2 Charakteristické vlastné čísla a podiely rozptylov hlavných komponentov

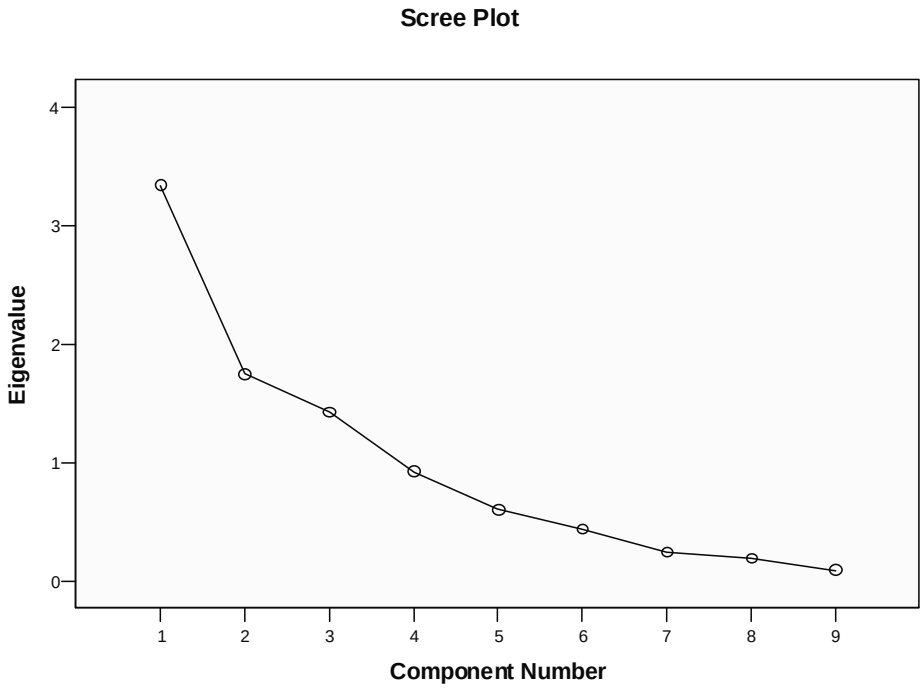
Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,338	37,090	37,090	3,338	37,090	37,090
2	1,745	19,385	56,475	1,745	19,385	56,475
3	1,424	15,820	72,295	1,424	15,820	72,295
4	,925	10,276	82,571	,925	10,276	82,571
5	,601	6,675	89,246	,601	6,675	89,246
6	,440	4,885	94,131	,440	4,885	94,131
7	,245	2,717	96,848			
8	,193	2,139	98,987			
9	,091	1,013	100,000			

Extraction Method: Principal Component Analysis.

- pomocou pravidla o minimálnej strate informácií, ktoré sú obsiahnuté v pôvodných premenných by sme sa rozhodovali o potrebe 5 respektíve 6 hlavných komponentov. V prípade, že vyberieme 5 prvých hlavných komponentov dôjde k vyčerpaniu 89,2% celkového rozptylu pôvodných premenných a ak by sme sa rozhodli, že vyberieme 6 prvých hlavných komponentov vyčerpáme nimi 94,1% celkového rozptylu pôvodných deviatich premenných,
- výber hlavných komponentov môžeme uskutočniť i pomocou *grafu 1*, ktorý zachytáva jednotlivé charakteristické vlastné čísla. Pomocou tohto grafu zvolíme taký počet prvých hlavných komponentov, pri ktorých dochádza k výraznému zlomu, pretože nevýrazné hlavné komponenty graficky predstavujú akési vodorovné dno. Výrazné hlavné komponenty sú teda oddelené zreteľným zlomovým miestom a hodnota indexu v bode zlomu na x-ovej osi udáva počet významných hlavných komponentov. Pomocou tohto grafu by sme určili 5 prvých hlavných komponentov.

Graf 1 Graf charakteristických vlastných čísel korelačnej matice



Po stanovení počtu významných hlavných komponentov prepočítame hodnoty pôvodných premenných pomocou koeficientov v_{jk}, ktoré sú prvkami transformačnej matice uvedenej v tabuľke 3.

Týmto výpočtom dôjde k stanoveniu nových, nemerateľných, latentných nezávislých premenných, teda k stanoveniu tzv. komponentného skóre.

Tabuľka 3 Transformačná matica

Component Score Coefficient Matrix

	Component							
	1	2	3	4	5	6	7	8
Zscore(A1)	-,201	,119	,388	,152	,432	,509	,348	-1,152
Zscore(A2)	,239	-,200	,213	-,011	,363	-,451	-,608	,230
Zscore(A3)	,085	,308	-,152	,789	-,121	-,093	-,808	-,394
Zscore(A4)	,028	,521	,021	-,141	-,117	-,625	,951	,352
Zscore(A5)	,164	-,202	-,401	,281	,497	-,037	1,179	-,502
Zscore(A6)	,204	,186	,334	-,025	,681	,011	-,049	,359
Zscore(A7)	,243	,092	,046	,041	-,333	1,125	,331	,667
Zscore(A8)	-,234	-,124	,095	,488	,237	-,029	,330	1,601
Zscore(A9)	-,107	,231	-,447	-,312	,664	,380	-,657	,384

Extraction Method: Principal Component Analysis.

Nové premenné, teda hodnoty hlavných komponentov je možné následne používať v analýzach namiesto pôvodných premenných, čím dôjde jednak k zmenšeniu dimenzie riešeného problému ako i k odstráneniu problému so silne korelovanými premennými.

Pomocou komponentných váh, teda pomocou korelačných koeficientov medzi j-tou pôvodnou premennou a k-tym hlavným komponentom (viď tabuľka 4), môžeme interpretovať jednotlivé hlavné komponenty.

Pri interpretácii hlavných komponentov sa obmedzujeme iba na komponenty s vysokými charakteristickými číslami, pretože interpretácia hlavných komponentov býva často obtiažna a problematická.

Tabuľka 4 Komponentné váhy – koeficienty korelácie medzi j-tou premennou a k-tym hlavným komponentom

Component Matrix[a]

	Component							
	1	2	3	4	5	6	7	8
Zscore(A1)	-,670	,207	,553	,141	,260	,224	,085	-,222
Zscore(A2)	,799	-,348	,303	-,010	,218	-,198	-,149	,044
Zscore(A3)	,283	,538	-,217	,729	-,073	-,041	-,197	-,076
Zscore(A4)	,093	,910	,030	-,130	-,070	-,275	,233	,068
Zscore(A5)	,548	-,353	-,571	,260	,298	-,016	,288	-,097
Zscore(A6)	,682	,324	,475	-,023	,409	,005	-,012	,069
Zscore(A7)	,810	,160	,065	,038	-,200	,495	,081	,128
Zscore(A8)	-,781	-,216	,135	,451	,142	-,013	,081	,308
Zscore(A9)	-,359	,403	-,636	-,288	,399	,167	-,161	,074

Extraction Method: Principal Component Analysis.

a. 8 components extracted.

Na základe tabuľky 4 môžeme skonštatovať, že prvý hlavný komponent je silne korelovaný s koeficientom rastu HDP, koeficientom rastu celkového počtu zamestnaných ako i s tvorbou hrubého fixného kapitálu. Tento prvý hlavný komponent teda môžeme popísať ako komponent zachytávajúci ekonomický rozmach jednotlivých krajín.

Druhý hlavný komponent je silne korelovaný s deflátorom HDP a harmonizovaným indexom spotrebiteľských cien, tento komponent teda zachytáva cenovú stabilitu jednotlivých sledovaných krajín.

Tretí komponent je najsilnejšie korelovaný s hodnotou HDP na obyvateľa v PKS. Popis ďalších hlavných komponentov je zložitejší, z tabuľky je zrejmé, že postupne jednotlivé koeficienty korelácie klesajú, teda vzťah medzi pôvodnými premennými a hlavnými komponentami nie je tak výrazný.

Záver

Analýzu hlavných komponentov odporúčame uskutočniť v prípade potreby redukcie dimenzie pôvodného súboru premenných, pričom dôjde k minimálnej strate informácií.

Ďalším cieľom, ktorý sledujeme analýzou hlavných komponentov je odstránenie silnej korelácie medzi pôvodnými premennými. Hlavné komponenty sú vzájomne nezávislé a preto vhodné pri použití rôznych analytických techník viacrozmernej štatistiky, ktoré predpokladajú vzájomnú nezávislosť analyzovaných premenných. V dôsledku slabej až strednej silnej korelácie medzi premennými, kritériami sa v našom konkrétnom príklade deväť rozmerný priestor podarilo znížil iba na päť rozmerný, nedošlo teda k výraznej redukcii dimenzie riešeného problému.

Článok bol spracovaný ako súčasť riešenia projektu VEGA 1/2555/05.

Literatúra:

[1] MELOUN, M. – MILITKÝ, J.: Statistická analýza experimentálních dat, Praha, Academia, 2004

[2] OSTERLIND, S.J. – TABACHNICK, B.G.: SPSS for Windows Workbook to accompany Using Multivariate Statistics, Needham Heights, Allyn&Bacon, 2001

www.spss.com, www.worldbank.org, www.europa.eu.int

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

megyesiova@euke.sk

Špecifické problémy sezónnej analýzy časových radov

Helena Glaser – Opitzová

Abstract

This paper deals with specific problems of seasonal adjustment of time series, with the risks of fully automatized decomposition of time series and looks at ways of quality improvement, and in particular comparability and harmonisation of seasonally adjusted figures.

Význam dekompozície časových radov netreba zvlášť zdôrazňovať. Na informáciách, ktoré vyplývajú zo sezónne upravených časových radov alebo radov trendu, sú často založené rozhodnutia hospodárskej politiky, lebo pôvodné údaje nie sú vhodným indikátorom krátkodobého vývoja ekonomiky. Harmonizované a porovnateľné sezónne upravené časové rady a rady trendu sú preto nevyhnutným predpokladom pre porovnanie tendencií krátkodobého vývoja ekonomík rôznych krajín. Na rozklad časových radov boli vytvorené rôzne metódy sezónnych úprav, ktoré rozkladajú predupravený časový rad do trendovo-cyklickej, sezónnej a nepravidelnej zložky.

$$X_t = T_t + S_t + I_t$$

kde T_t je trendovo-cyklická zložka, S_t je sezónnosť a I_t je nepravidelná zložka. (Z dôvodu zjednodušenia interpretácie neuvažujeme ani transformáciu údajov a neberieme do úvahy ani ďalšie zložky časového radu, ako napr. kalendárne variácie).

Čo je trend a čo je sezónne upravený časový rad

Ekonomické teórie neposkytujú jednoznačnú definíciu zložiek časových radov.

Zložky sú zvyčajne definované nasledovne:

Trend časového radu charakterizuje dlhodobú tendenciu časového radu rásť alebo klesať, teda jeho základnú vývojovú tendenciu. Trend chápeme ako výsledok systematicky pôsobiacich činiteľov.

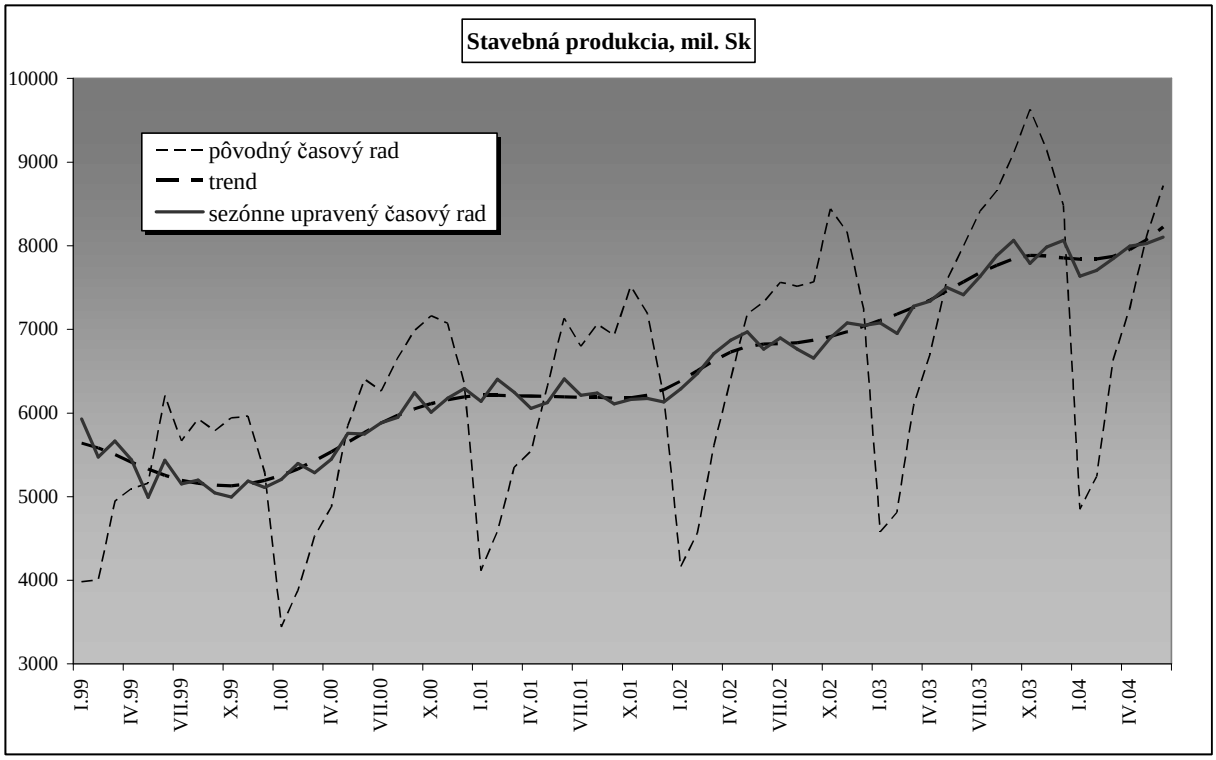
Sezónna zložka predstavuje viac či menej pravidelne sa opakujúce kolísanie hodnôt časového radu okolo trendu, pričom dĺžka vlny sa rovná jednému roku. Tieto výkyvy sú spôsobené rôznymi činiteľmi súvisiacimi s klimatickými podmienkami, so striedaním ročných období

a rozdielmi v teplote počas roka a so všetkými ich dôsledkami na ekonomické a spoločenské aktivity. Príčinou sezónneho kolísania v časových radoch môžu byť i určité tradície a zvyklosti, ale môže ho vyvolávať aj metodika vyhodnocovania hospodárenia podnikov (výplaty štvrťročných a ročných prémií) a podobne.

Cyklická zložka predstavuje periodický výkyv údajov časového radu okolo trendu, s dĺžkou cyklu väčšou ako rok, pričom dĺžka cyklu, t. j. vzdialenosť medzi dvoma susednými bodmi zvratu, sa často mení. Cyklické kolísanie vzniká v dôsledku vedecko – technických a inovačných procesov, alebo v dôsledku zmenenej hospodárskej politiky štátu, či príslušnej organizácie.

Treba poznamenať, že vzhľadom na problémy vznikajúce pri separácii trendovej a cyklickej zložky , niektorí autori cyklickú zložku nechápu ako samostatnú zložku časového radu, ale trendová a cyklická zložka sa často poníma ako kombinovaná ***cyklicko-trendová zložka*.**

Nakoniec treba uvažovať s nepravidelným kolísaním časového radu, ktoré sa prisudzuje pôsobeniu drobných, často nepostihnuteľných a nepredvídateľných, vzájomne nezávislých príčin alebo je dôsledkom náhodne sa vyskytujúcich udalostí. Tento zvyšok, ktorý zostáva v časovom rade po vylúčení trendovej a periodickej (sezónnej a cyklickej) zložky a o ktorom predpokladáme, že má náhodný charakter, sa nazýva **náhodná zložka**.



Predpokladáme, že tieto zložky sú nezávislé, aj keď praktické skúsenosti tomuto predpokladu niekedy protirečia.

Dobrým príkladom je stavebný priemysel. Pozorujeme tu sezónny efekt zimného obdobia (teplota klesá pod 4 stupne Celzia), ktorý je závislý od aktuálnych okolností business cyklu. Ak sa stavebníctvo nachádza v recesii, produkcia sa zastavuje. Ak je stavebníctvo v období konjunktúry, do betónu sa pridávajú chemikálie a produkcia pokračuje aj napriek nízkym teplotám.

Pomerne nepresné definície jednotlivých zložiek časového radu vedú k dvom základným problémom:

1. neschopnosť vedcov vysvetliť, prečo robíme rozklad časových radov,
2. množstvo metód používaných na sezónnu úpravu s rozdielnymi definíciami zložiek časového radu.

Jasnú a zreteľnú odpoveď na prvý problém – prečo aplikovať sezónnu úpravu – dali Bell a Hillmer (1992): *Sezónnu úpravu aplikujeme za účelom zjednodušenia údajov, aby ich aj štatisticky neskúsený užívateľ mohol oveľa ľahšie interpretovať bez podstatnej straty informácií.*

Čo sa týka druhého problému, na ilustráciu rozdielnych definícií zložiek časových radov v rôznych procedúrach by bolo možné skúmať rôzne metódy, ako napr. BV4, X-11 a SEATS vo frekvenčnej oblasti a porovnať zisky jednotlivých filtrov. Skúmať, či rôzne frekvencie zložiek časového radu, keď prejdú daným filtrom sú nezmenené, alebo úplne odstránené, čo predstavuje samozrejme hraničné prípady.

Ako už bolo uvedené, definície zložiek časového radu sa od seba významne líšia. Okrem toho obe zložky, či už **trend alebo sezónne upravený časový rad**, ktoré publikujú štatistické úrady, **neposkytujú presné hodnotenia alebo výsledky, ale ponúkajú rady a pomoc** riadiacim subjektom pri krátkodobom monitoringu ekonomiky. Navzdory diskusiám o prednostiach či už trendu alebo sezónne upraveného časového radu treba mať tieto skutočnosti na pamäti.

Taktiež by mali byť zohľadnené dva rozdielne pohľady, a to pohľad retrospektívny a perspektívny.

Z retrospektívneho hľadiska je cieľom rozkladu časového radu ponúknuť čo najpresnejšie informácie o vývoji ekonomiky. Ak je nepravidelná zložka časového radu relatívne významná,

tak v sezónne upravenom časovom rade môže byť táto informácia ťažšie identifikovateľná a trend nám ponúka lepší nástroj pre analýzu.

Z perspektívneho hľadiska je pre nás dôležitý aktuálny koniec časového radu. Pretože na konci časového radu nie je možné použiť pre rozklad symetrické filtre, je treba použiť asymetrické filtre alebo aplikovať symetrické po predĺžení časového radu prognózovaním. Oba prístupy vedú k revíziám a ich obraz je odlišný pre trend a pre sezónne upravený časový rad.

Trend na konci časového radu má sklon k väčším revíziám ako sezónne upravený časový rad a teda je jednoduchšie odhadnúť relatívne stabilnú sezónnu zložku ako trend.
Ide tu o konflikt medzi získaním čo najpresnejšej informácie o vývoji ekonomiky a požiadavkou čo najmenších revízií.

Z tohto dôvodu sa väčšina analytikov rozhoduje publikovať na konci časového radu sezónne upravené údaje. Avšak okrem revízií tu máme ešte problém so ziskom čo najpresnejších informácií o vývoji. Jednoduché porovnanie revízií medzi trendom a sezónne upraveným časovým radom nie je preto postačujúce. Okrem toho musíme nájsť aj spôsob, ako získať čo najpresnejšie informácie zo sezónne upraveného časového radu, ktorý obsahuje nepravidelnú zložku. Teoreticky, je to zložka časového radu, ktorá obsahuje cenné informácie pre analýzy, ale len v tom prípade, ak má analytik rozsiahle doplňujúce znalosti o detailoch v ekonomike a presne pozná predpoklady procedúry sezónnej úpravy.

Celkovo možno konštatovať, že pre analýzu historických dát je vhodné použiť trend-cyklus, ale sporný zostáva koniec časového radu.

Konflikt existuje aj medzi rozsahom revízií a rýchlosťou konvergencie smerom k stabilným výsledkom. Dá sa dokázať teoreticky aj empiricky, že trend-cyklus vykazuje počas niekoľkých mesiacov podstatne viac revízií ako sezónne upravený časový rad, ale na druhej strane sa hodnoty omnoho rýchlejšie stabilizujú do finálnej podoby business cyklu, a tak odhaľujú tento obraz jasnejšie. V 95 % prípadov nám trend-cyklus dáva jasnejší pohľad na ekonomickú situáciu a pomáha užívateľom lepšie porozumieť tomu, či ekonomika rastie alebo klesá.

Publikovanie trend-cyklu a sezónne upraveného časového radu by sa nemalo navzájom vylučovať. Prvý menovaný je vhodnejší pre laikov. Ale odborníci potrebujú pre svoje ekonomické analýzy rozsiahly súbor informácií, vrátane sezónne upravených časových radov.

Ako už bolo uvedené, metódy sezónnych úprav rozkladajú **predupravený časový rad**. S predúpravou časových radov sú spojené také pojmy, ako kalendárna zložka časového radu (korekcia na kalendárne variácie) alebo extrémne výkyvy.

Kalendárna zložka

Kalendárna zložka, zahrnuje diferencie v štruktúre kalendára, napríklad rozdielny počet pracovných dní v rámci jednotlivých mesiacov alebo špeciálne efekty, ako napr. efekt Veľkej noci. Odhad kalendárnej zložky je obzvlášť dôležitý v prípade tokových premenných. Kalendárna zložka je odhadovaná regresnými metódami a absencia tejto predúpravy časového radu môže spôsobiť narušenie stability sezónnosti.

Odlišný počet pracovných dní v jednotlivých mesiacoch alebo štvrťrokoch môže ovplyvniť úroveň ekonomickej činnosti v špecifikovanom období. Korekcia tokových radov údajov vzhľadom na pracovné/obchodné dni je dôležitým krokom v krátkodobých analýzach časových radov. Rôzny počet týchto dní v danom kalendárnom období môže ovplyvniť časové rady niekoľkými percentuálnymi bodmi. Rôzny výskyt prázdnin a iných príčin absencie v práci v rámci kalendárneho obdobia môže mať ten istý efekt. Iba korekcia týchto efektov poskytuje možnosť vytvorenia zmysluplných porovnaní daného štvrťroka s jeho náprotivkom v predchádzajúcich rokoch. Keď sú porovnania rozšírené z národnej úrovne na medzinárodnú, je znalosť týchto efektov ešte významnejšia.

Správne korekcie týchto efektov môžeme urobiť iba na základe detailných znalostí kalendára a ďalších charakteristík, ktoré sa týkajú týchto dní v jednotlivých krajinách.

Extrémne výkyvy

Extrémne výkyvy ovplyvňujú kvalitu prognózovania, trendu a sezónne upraveného časového radu tým, že spôsobujú vychýlenie modelu používaného pre rozklad časového radu.
Časové rady, ak je to možné, kontrolujeme na prítomnosť extrémnych hodnôt na čo najnižšej úrovni agregácie, pretože v agregovanom časovom rade sa môžu strácať a táto skutočnosť môže viesť k chybnej interpretácii aktuálnej situácie.

Vyhľadávané a upravované sú štyri typy extrémnych hodnôt (inovácie, aditívne extrémne výkyvy, zlomy trendu a prechodné zmeny).

Stupeň automatizácie rozkladu časových radov

Neexistuje jednoznačný názor do akej miery je nutný zásah užívateľa do programu na sezónnu úpravu, ale v každom prípade je dôležité mať prehľad o možných rizikách zlého výberu modelu, ktorý spôsobí napr. odhalenie nezmyselných extrémnych hodnôt alebo absurdné výsledky.

Aj u takých prepracovaných programov ako je napríklad TRAMO/SEATS sa môže vyskytnúť vážny problém. Predupravujúci program TRAMO vyberie sezónny model pre časový rad s veľmi malou alebo žiadnou sezónnosťou a SEATS tento model zmení na iný sezónny model a použije ho na nesezónny časový rad. Takýmto spôsobom môže zaviesť sezónnosť do sezónnej úpravy nesezónneho časového radu. Zabrániť tomuto problému a odhaliť časové rady, ktoré by nemali byť upravované, môže použitie spektrálnej diagnostiky, a teda aktívny prístup užívateľa.

Závery sú teda jednoznačné. Dokonca aj v tých najsofistikovanejších procedúrach sezónnych úprav je zásah analytikov nevyhnutný.

Toto sú len niektoré špecifické problémy sezónnej analýzy časových radov. Ale aj riešenie čiastkových problémov môže viesť k zlepšeniu kvality, a najmä k porovnateľnosti a harmonizácii sezónne upravených hodnôt.

Použitá literatúra:

1. Eurostat Recommendations concerning Seasonal Adjustment Policy, EUROSTAT 1996-1998
2. Seasonal Adjustment Methods, A Comparison for Industry Statistics, EUROSTAT, October 1998
3. Comparison of Time Series Characteristics for Seasonal Adjustment from SEATS and X12-ARIMA, Catherine C. Hood, U.S. Census Bureau, Washington

Ing. Helena Glaser – Opitzová, INFOSTAT, Dúbravská cesta 3, 845 24 Bratislava, glaser@infostat.sk

Metódy na zaistenie konzistencie štvrťročných a ročných časových radov. (Benchmarking)

Helena Solčánska

Abstract

Both, the quarterly and annual estimates of GDP are compiled differently by using different data sources. It is necessary to re-align quarterly estimates with the annual benchmarks in order to ensure consistency. This paper shortly describes two methods for assurance this consistence of time series.

Úvod

Vzhľadom na to, že štvrťročné a ročné odhady tvorby a použitia HDP sú vytvorené na základe rozdielnych zdrojov dát, pri ich zostavovaní je potrebné zabezpečiť, aby boli navzájom konzistentné. To znamená, aby súčet štyroch štvrťročných hodnôt zodpovedal ročnej hodnote. Postup, ktorý takúto konzistenciu zabezpečuje sa nazýva benchmarking. Jednoduchšie metódy, ktoré sa využívajú na riešenie takejto úlohy spôsobujú tzv. step problém. Preto boli vyvinuté viaceré metódy, ktoré nespôsobujú „step problém“ a zároveň zaručujú konzistenciu štvrťročných a ročných časových radov. Jednou z nich je aj proporcionálna Dentonova metóda, ktorá zabezpečuje, že upravené štvrťročné hodnoty sú v maximálne možnej miere proporcionálne s pôvodnými hodnotami. Zároveň sa táto metóda využíva na predbežné odhady ukazovateľov v tých prípadoch, keď ešte nie sú známe ich ročné hodnoty.

Čo je benchmarking

Pri zostavovaní odhadov tvorby a použitia HDP na základe metodiky štvrťročných a ročných národných účtov je potrebné zaistiť súlad medzi štvrťročnými časovými radmi a zodpovedajúcimi ročnými radmi. Všeobecne je prijatý predpoklad, že ročné údaje sú spoľahlivejšie, a preto je potrebné upraviť štvrťročné údaje tak, aby pre každý rok platila rovnosť medzi súčtom štvrťročných hodnôt a príslušnou ročnou hodnotou. Postup, pri ktorom je zaistená táto konzistencia (tzv. benchmarking), spočíva v dvoch krokoch:

a, dezagregácia, ktorej účelom je rozdeliť ročné hodnoty do štvrťrokov na základe využitia pôvodných štvrťročných údajov

b, extrapolácia, ktorej cieľom je predĺžiť existujúci časový rad tak, aby si neskôr, na základe známej ročnej hodnoty, vyžadoval len minimálnu revíziu.

Uvedené dva kroky benchmarkingu sa navzájom nepodmieňujú, čo znamená, že ich možno vykonávať nezávisle jeden od druhého.

Benchmarking sa najčastejšie uplatňuje na časové rady so štvrťročnou a ročnou periodicitou, ale v zásade je možné rovnaký postup použiť na časové rady s ľubovoľnými periodicitami (napr. mesačné a ročné, mesačné a štvrťročné, ročné a viacročné).

Cieľom benchmarkingu je zachovať, v maximálne možnej miere vývoj sledovaného ukazovateľa daného pôvodnými štvrťročnými údajmi a pritom rešpektovať podmienku súladu s ročným údajom. Jednoduché metódy pre zaistenie takejto konzistencie, spočívajúce v rozvrhnutí rozdielu medzi súčtom štvrťrokov a ročnou hodnotou proporcionálne k veľkostiam pôvodných hodnôt, túto požiadavku nespĺňajú. Spôsobujú totiž neopodstatnené skoky v údajoch za jednotlivé roky (tzv. step problém). Jedna z nich, ktorá je známa ako Pro-rata rozdelenie, je charakterizovaná a ilustrovaná v nasledujúcej časti.

Pro–rata rozdelenie a step problém

Najjednoduchšia metóda na získanie štvrťročného odhadu z príslušného ročného údaju je Pro-rata rozdelenie. Matematický vzťah na výpočet štvrťročného odhadu pomocou Pro-rata rozdelenia je:

$$X_{q,T} = A_T \left(\frac{I_{q,T}}{\sum_q I_{q,T}} \right) = A_T \, ABI$$

kde

$X_{q,T}$ je hodnota štvrťročného odhadu pre štvrťrok q roku T

$I_{q,T}$ je hodnota pôvodného ukazovateľa v štvrťroku q roku T

A_T je hodnota ročného benchmarkingového údaju pre rok T

ABI annual benchmark-to-indicator ratio, pomer ročnej hodnoty a súčtu pôvodných štvrťročných hodnôt

Výsledky získané touto metódou sú však problematické. Indikátor ABI nie je totiž stabilný, ale sa mení z roka na rok, pretože pôvodný štvrťročný ukazovateľ sa (z celoročného hľadiska) vyvíja v porovnaní s ročným ukazovateľom spravidla rozdielne. V prípade Pro-rata

rozdelenia navyše platí, že indikátor ABI je zhodný s indikátorom QBI (quarterly benchmark to-indicator-ratio), a to pre každé obdobie (v našom prípade štvrťrok) v danom roku. V konečnom dôsledku sú teda QBI pre susedné roky rozdielne, v dôsledku čoho vyúsťuje úprava štvrťročných ukazovateľov pomocou tejto metódy k tzv. step problému.

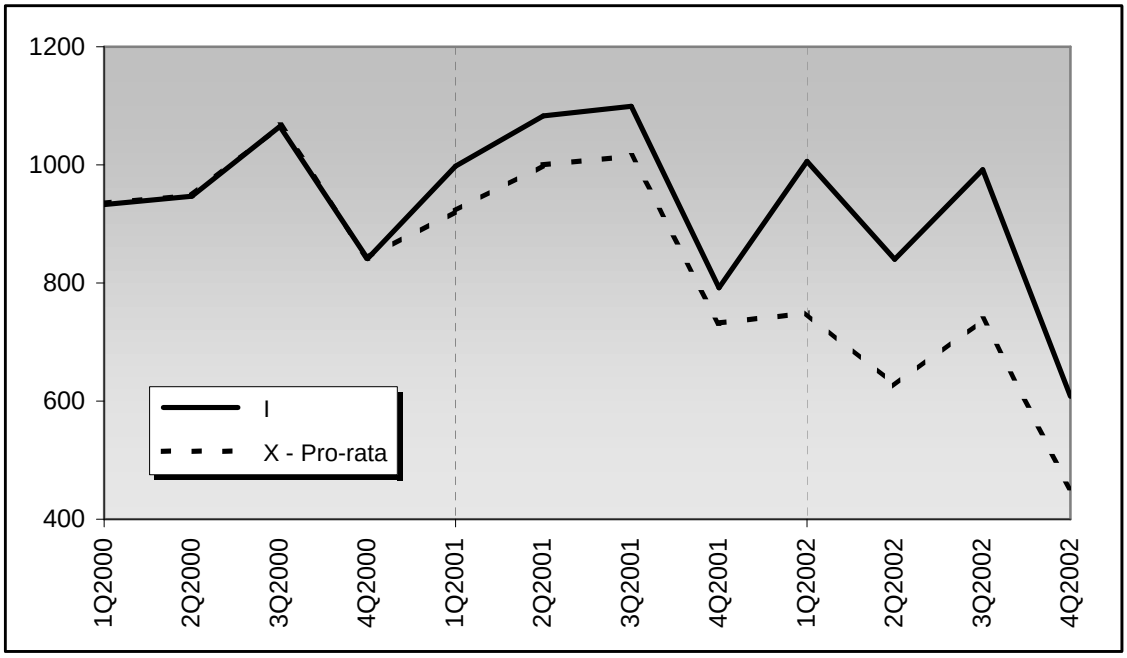
Názornú ilustráciu tohto problému poskytuje tab.1 a graf 1.

Údaje v tab. 1 ilustrujú rozklad ročnej hodnoty ukazovateľa (ročný benchmark) do štvrťrokov pomocou Pro-rata rozdelenia s využitím pôvodných štvrťročných hodnôt daného ukazovateľa. Podstatu tzv. step problému môžeme identifikovať z porovnania stĺpcov „ukazovateľ“ a „štvrťročný odhad“ medzištvrťročných relatívnych zmien v tab.1. Vzhľadom na to, že QBI sú pre susedné roky rozdielne, použitie tejto metódy spôsobí zmenu v relatívnych medzištvrťročných zmenách len v 1.štvrťroku daného roku, pričom ostatné relatívne medzištvrťročné zmeny zostávajú nezmenené. Úprava štvrťročných ukazovateľov prostredníctvom tejto metódy spôsobuje neopodstatnené skoky v ich vývoji. Napr. v 1. štvrťroku 2002 došlo k značnej modifikácii medzištvrťročnej relatívnej zmeny z 27% na 2.4%. Súvisí to najmä s výrazne rozdielnym poklesom pôvodného štvrťročného a ročného (benchmark) ukazovateľa v roku 2002. Kým prvý z nich oproti roku 2001 poklesol o 13.2%, druhý až o 30.1%. Z grafu 1 je tiež zrejmé, že pôvodný štvrťročný ukazovateľ a upravený štvrťročný ukazovateľ majú odlišný vývoj.

Tab. 1 ***Pro-rata rozdelenie***

obdobie	ukazovateľ		ročný benchmark		ročný BI	štvrťročný odhad		štvrťročný BI
	I (mil. Sk) (1)	relatívne medzištvrťročné zmeny	A (2)	relatívne medziročné zmeny	ABI (2)/(1)	X (4)	relatívne medzištvrťročné zmeny	QBI (4)/(1)
1Q2000	933	-23.3				935	**-11.6**	1.002
2Q2000	947	1.4				948	1.4	1.002
3Q2000	1065	12.5				1067	12.5	1.002
4Q2000	841	-21.0				843	-21.0	1.002
2000	**3786**	**-16.8**	**3792**	**-4.0**	**1.002**	**3792**		**1.002**
1Q2001	998	18.7				922	**9.4**	0.923
2Q2001	1083	8.5				1000	8.5	0.923
3Q2001	1099	1.4				1015	1.4	0.923
4Q2001	792	-27.9				732	-27.9	0.923
2001	**3973**	**4.9**	**3669**	**-3.3**	**0.923**	**3669**		**0.923**
1Q2002	1006	27.0				749	**2.4**	0.744
2Q2002	840	-16.5				626	-16.5	0.744
3Q2002	992	18.1				739	18.1	0.744
4Q2002	608	-38.8				452	-38.8	0.744
2002	**3447**	**-13.2**	**2566**	**-30.1**	**0.744**	**2566**		**0.744**

Graf 1.



Proporcionálna Dentonova metóda

V súčasnosti existujú viaceré metódy na zaistenie konzistencie medzi štvrťročnými a ročnými časovými radmi, ktoré nespôsobujú „step problém“. Ich prehľad uvádza manuál MMF k štvrťročným národným účtom [1]. Jednou z týchto metód je proporcionálna Dentonova metóda

Podstata Dentonovej metódy spočíva v tom, aby štvrťročné hodnoty boli upravené v maximálne možnej miere proporcionálne s pôvodnými hodnotami. To sa dosahuje riešením optimalizačnej úlohy, ktorej matematické vyjadrenie je nasledovné:

$$\min \sum_{t=2}^{N} \left[\frac{X_t}{I_t} - \frac{X_{t-1}}{I_{t-1}} \right]^2$$

za podmienky, že platí obmedzenie

$$\sum_{t=4y-3}^{4y} X_t = A_y, \qquad y = 1, ..., T,$$

To znamená, že suma štvrťrokov pre každý rok sa rovná príslušnému ročnému údaju.

kde

t označuje čas ($t = 4y$-3 je prvý štvrťrok roku y, $t = 4y$ je štvrtý štvrťrok roku y);

X_t je hodnota štvrťročného odhadu pre štvrťrok t;

I_t je pôvodná hodnota ukazovateľa pre štvrťrok t;

A_y je hodnota ročného benchmarkingového údaju pre rok y;

T je posledný rok, pre ktorý máme k dispozícii ročnú hodnotu;

N je posledný štvrťrok, pre ktorý máme k dispozícii štvrťročné zdroje dát

Numerická ilustrácia predchádzajúceho príkladu použitím Dentonovej metódy je uvedená v tab. 2.
Jej posledný stĺpec obsahuje vypočítané štvrťročné QBI ktoré sa zo štvrťroka na štvrťrok menia postupne, na rozdiel od Pro-rata rozdelenia, kde sa QBI menili náhle z roka na rok. V prípade použitia Dentonovej metódy sa QBI v rámci roka menia, ale priemerné ABI zostávajú nezmenené.
Navyše si môžeme všimnúť, že všetky relatívne medzištvrťročné zmeny sú v porovnaní s pro-rata rozdelením rozdielne, ale ich priebeh je plynulejší.

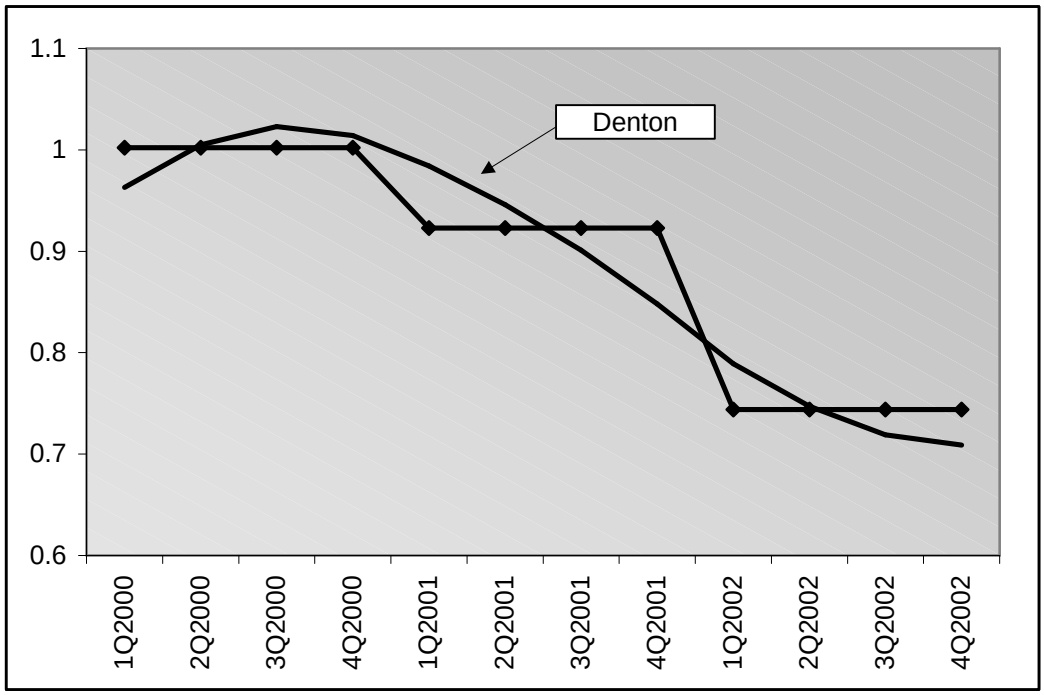
Priebeh QBI získaných Pro-rata rozdelením a Dentonovou metódou je zobrazený na grafe 2.

Graf. 3 ilustruje porovnanie štvrťročných časových radov pôvodného ukazovateľa, odhadov získaných Pro–rata rozdelením a Dentonovou metódou.

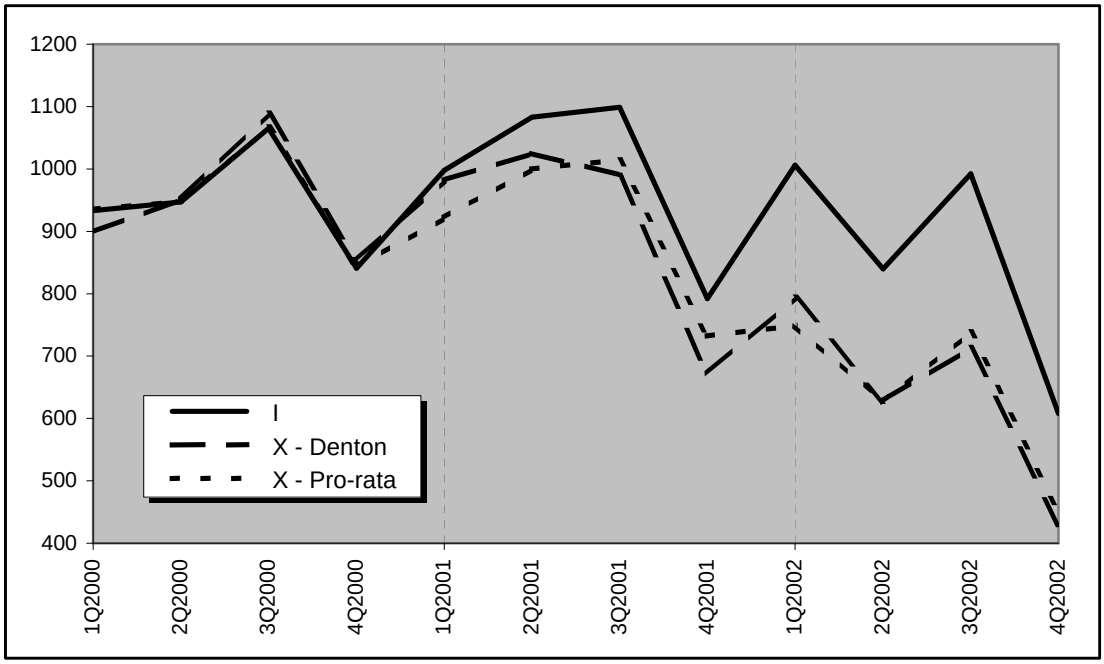
Tab.2 ***Proporcionálna Dentonova metóda.***

obdobie	ukazovateľ		ročný benchmark		ročný BI	štvrťročný odhad		štvrťročný BI
	I (mil. Sk) (1)	relatívne medzištvrťročné zmeny	A (2)	relatívne medziročné zmeny	ABI (2)/(1)	X (4)	relatívne medzištvrťročné zmeny	QBI (4)/(1)
1Q2000	933	-23.3				899	-17.7	0.963
2Q2000	947	1.4				951	5.8	1.005
3Q2000	1065	12.5				1089	14.5	1.024
4Q2000	841	-21.0				853	-21.7	1.015
2000	**3786**	**-16.8**	**3792**	**-4.0**	**1.002**	**3792**		**1.002**
1Q2001	998	18.7				982	15.1	0.985
2Q2001	1083	8.5				1025	4.4	0.949
3Q2001	1099	1.4				990	-3.4	0.907
4Q2001	792	-27.9				672	-32.2	0.849
2001	**3973**	**4.9**	**3669**	**-3.3**	**0.923**	**3669**		**0.923**
1Q2002	1006	27.0				794	18.1	0.792
2Q2002	840	-16.5				628	-20.9	0.749
3Q2002	992	18.1				714	13.7	0.722
4Q2002	608	-38.8				430	-39.7	0.709
2002	**3447**	**-13.2**	**2566**	**-30.1**	**0.744**	**2566**		**0.744**

Graf 2.



Graf 3.



Záver

Prezentované výsledky použitia proporcionálnej Dentonovej metódy boli získané pomocou špecializovaného softwaru ECOTRIM. Je to software vyvinutý Eurostatom, slúžiaci ako pomôcka pre zabezpečenie konzistencie časových radov, ktorý je schopný prepájať sa s príbuznými aplikáciami a balíkmi. Okrem Ecotrim-u existuje viacero špecializovaných softwarov pre praktickú realizáciu benchmarkingu napr. Bench, Flash.

Samozrejme, že existuje veľa metód a prístupov na zaistenie konzistencie medzi štvrťročnými a ročnými časovými radmi, ale dôkladná analýza a pochopenie tejto problematiky si vyžaduje rozsiahlejšiu štúdiu a väčšie možnosti z hľadiska softwarového vybavenia.

Použitá literatúra:

1. Quarterly National Accounts Manual - Concepts, Data Sources, and Compilation, International Monetary Fund 2001
2. Benchmarking and Interpolation, ECD/ADB/ESCAP/ Workshop on Quarterly national Accounts, OECD 2002
3. Statistical Benchmarking, Algera Symon, OECD/ Eurostat workshop, April 2005

Ing. Helena Solčánska, INFOSTAT, Dúbravská cesta 3, 845 24 Bratislava, solcanska@infostat.sk

Analýza kyvadlovej migrácie v okresoch severovýchodného Slovenska

Alena Kaščáková, Gabriela Nedelová

Abstract

The aim of this paper is to describe and analyze the situation of the migration in selected region of the north-eastern Slovakia. The survey by means of questionnaire enabled to describe daily or short-term migration specially of men and women in chosen region. The regression analysis of alternative variable presents the significant variables influencing the selected type of the migration The logit model is constructed.

Úvod

Zmeny v počte a štruktúre obyvateľstva sú vyvolávané reprodukciou obyvateľstva i jeho mobilitou. V súvislosti s týmito zmenami definuje odborná demografická literatúra prvý druh pohybu ako prirodzený pohyb, druhý označuje ako priestorový alebo mechanický pohyb obyvateľstva. Mobilita obyvateľstva zahŕňa všetky druhy priestorového pohybu obyvateľstva, trvalý aj prechodný pohyb, pohyb z rôznych príčin, smerov i vzdialenosti, s rôznou periodicitou aj úrovňou organizovanosti.

Vzhľadom k tomu, že v poslednom období výrazne vzrástol rozsah zahraničnej migrácie, pozornosť demografickej ale aj laickej verejnosti sa upiera k fenoménu migrácie. V súvislosti so zmenami politického aj ekonomického systému a sociálnych podmienok života obyvateľstva je potrebné zamerať svoju pozornosť nielen na úroveň zahraničnej, ale aj vnútornej migrácie. V rámci nej hrá významnú úlohu vnútorná migrácia za prácou a to trvalá aj dočasná (kyvadlová). V prípade kyvadlovej migrácie ide o pravidelnú dennú či týždennú dochádzku do zamestnania, teda pravidelne sa opakujúci pohyb medzi bydliskom a pracoviskom.

Súbor sledovaných okresov

V predkladanom príspevku sme sa snažili modelovať situáciu v domácnostiach a vytipovať významné znaky ovplyvňujúce úroveň kyvadlovej migrácie v niektorých

okresoch severovýchodného Slovenska. Zvolili sme metódu dotazníkového prieskumu v súbore 328 domácností, ktoré pokryli priestor okresov Poprad, Kežmarok, Spišská Nová Ves, Prešov a Stará Ľubovňa. Ide o okresy s pomerne vyššou mierou nezamestnanosti a obmedzenými možnosťami získania zamestnania. V sledovaných domácnostiach sme zisťovali situáciu v dochádzke za prácou zvlášť u mužov a zvlášť u žien, ktorí odpovedali zvlášť na celkovo 11 otázok, odpovede na 4 otázky poskytovali sociodemografický obraz domácnosti.

Najpočetnejšie boli zastúpené štvorčlenné domácnosti (celkovo 148), druhé v poradí nasledovali trojčlenné domácnosti. V sledovaných domácnostiach je v priemere 1,32 závislého dieťaťa, v domácnostiach pracuje 96,7 % mužov, 87,6 % žien. Z celkového počtu mužov, dochádzajúcich za prácou 88,5 % dochádza denne, 3 % týždenne, zvyšok s nižšou periodicitou. U žien je situácia menej variabilná, 98,5 % dochádzajúcich žien cestuje za prácou denne. Dochádzajúci muži prevažne cestujú do práce autom (47 %), autobusom (26 %) a pešo (16 %), ženy najčastejšie autobusom (39 %), pešo (34 %) a autom (22 %). Po vylúčení extrémnych hodnôt zisťujeme, že muži cestujú v priemere 0,53 a ženy 0,18 hodiny, keď priemerná vzdialenosť, ktorú pri dochádzaní do zamestnania prekoná muž je 38 km a žena necelých 9 km. Náklady na dochádzanie do zamestnania sú adekvátne vyššie u mužov – 971 Sk v priemere mesačne, zatiaľ čo u žien 584 Sk na mesiac. Dochádzajúce ženy najčastejšie trávia čas rozhovorom so známymi, muži šoférovaním. Spôsob dochádzania do zamestnania rovnako mužom aj ženám v prevažnej miere úplne vyhovuje, prípadne nevadí. Tí dochádzajúci, ktorí na predchádzajúcu otázku odpovedali negatívne ako hlavný dôvod svojej nespokojnosti označili finančnú náročnosť cestovania za prácou (muži) a časovú náročnosť (ženy). So vzdialenosťou dochádzky do zamestnania klesá počet dochádzajúcich mužov aj žien, u žien je rýchlosť poklesu početnosti výraznejšia ako u mužov. Muži aj ženy dochádzajú za prácou prevažne z toho dôvodu, že nenašli adekvátnu prácu v mieste svojho bydliska alebo dochádzajú za lepšie finančne ohodnotenou prácou.

Model závislosti dochádzania do zamestnania

Znak dochádzanie do zamestnania je v dotazníku definovaný ako dichotomický štatistický znak s obmenami 0 – ak respondent za prácou nedochádza mimo obec bydliska a s obmenou 1 – inak. Na modelovanie dochádzania do zamestnania sme využili logistickú regresiu pre binárnu vysvetľovanú premennú. Do modelu sme ako vysvetľujúce faktory zahrnuli počet členov domácnosti, či je partner zamestnaný alebo nie, počet závislých detí

v rodine, veľkosť obce, vzdialenosť, čas (x_5) a náklady na cestovanie do zamestnania (x_7). Postupným vyraďovaním nevýznamných faktorov sme pre mužov odhadli šancu dochádzania do zamestnania modelom

$$\hat{\pi} = \frac{e^{(-2.935+5.843x_5+0.002x_7)}}{1+e^{(-2.935+5.843x_5+0.002x_7)}}.$$

Podiel modelom úspešne zaradených prípadov k celkovému počtu prípadov je 62,16 %. Štatistika Hosmerovho a Lemeshowho testu je nevýznamná, t.j. medzi zistenými a očakávanými početnosťami je dobrá zhoda (p-hodnota 0,813). Odhadnuté koeficienty s ich štandardnými odchýlkami a s Waldovou testovacou štatistikou sú uvedené v nasledujúcej tabuľke (systém SPSS13).

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 6(a)	O5M	5.843	1.058	30.479	1	.000	344.835	43.323	2744.735
	O7M	.002	.000	37.838	1	.000	1.002	1.002	1.003
	Constant	-2.935	.411	51.117	1	.000	.053		

a Variable(s) entered on step 1: O2Z, O13, O15, O1, O5M, O6M, O7M.

Keďže dotazníky boli v domácnostiach vypĺňané zvlášť mužmi aj ženami, analyzovali sme situáciu v dochádzaní do zamestnania aj u žien.

$$\hat{\pi} = \frac{e^{(-3.744+7.246x_5+0.003x_7)}}{1+e^{(-3.744+7.246x_5+0.003x_7)}}$$

Podiel modelom úspešne zaradených prípadov k celkovému počtu prípadov je 52,63 %. Štatistika Hosmerovho a Lemeshowho testu je nevýznamná (p-hodnota 0,609). Odhadnuté koeficienty s ich štandardnými odchýlkami a s Waldovou testovacou štatistikou sú uvedené v tabuľke:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 6(a)	O5Z	7.246	1.265	32.833	1	.000	1402.709	117.638	16725.864
	O7Z	.003	.000	32.931	1	.000	1.003	1.002	1.004
	Constant	-3.744	.464	65.238	1	.000	.024		

a Variable(s) entered on step 1: O13, O15, O1, O2M, O5Z, O6Z, O7Z.

Logistická regresia umožnila odhadnúť pravdepodobnosť, či muž (žena) sledovanej domácnosti dochádza do zamestnania mimo obce alebo nie a čím tento fakt spôsobený. V prípade oboch pohlaví sa prejavili ako významné len premenné čas dochádzania do zamestnania a výška nákladov na cestovanie. Počet členov domácností, počet závislých detí ani veľkosť obce sa neprejavili ako významné premenné ovplyvňujúce skutočnosť, či respondent dochádza do zamestnania alebo nie.

Použitá literatúra

1. Hebák, P., Hustopecký J., Pecáková, I., Průša, M., Řezanková, H., Svobodová, A., Vlach, P. 2005. Vícerozměrné statistické metody (3). Informatorium. Praha, 2005. ISBN 80-7333-039-3.
2. Roubíček, V. 1997. Úvod do demografie. CODEX Bohemia. Praha, 1997. ISBN 80-85963-43-4.
3. Luha, J. 1998. Štatistické skúmanie závislosti kvalitatívnych znakov. In: Aplikácie štatistických metód v transformujúcej sa ekonomike. SŠDS. Bratislava, 1998. ISBN 80-967658-2-5.
4. Pecáková, I. 2005. Testy nezávislosti v řídkých kontingenčních tabulkách. In: Forum Statisticum Slovacum. SŠDS. Bratislava, 2005. ISSN 1336-7420.
5. Luha, J. 2005. Reprezentatívnosť vo výskumoch verejnej mienky. In: Forum Statisticum Slovacum. SŠDS. Bratislava, 2005. ISSN 1336-7420.

Ing. Alena Kaščáková, PhD.
EF UMB Banská Bystrica
alena.kascakova@umb.sk

RNDr. Gabriela Nedelová, PhD.
EF UMB Banská Bystrica
gabriela.nedelova@umb.sk

VZDIALENOSTNÉ OPERÁTORY - GEOINFORMAČNÉ NÁSTROJE PRIESTOROVEJ ŠTATISTIKY

Dagmar Kusendová

Abstract

The progress in technologies of geographic information systems (GIS) widens the application potential for the modelling spatial or geographic structures. These technologies encompass different spatial (geo)statistical methods employed in GIS-oriented programs, e.g. MapInfo, ArcView. Certain number of them is represented by distant operators used for the creation and analysis of spatially arranged demogeographical objects.

The article describes the problems related to the GIS distant operators and their use in the context of modelling and analysis of spatially differentiated demogeographic structures. Methodical aspects of spatial analyses and composition of geographic statistical fields (surfaces) using Thiessen polygons (Voronoi/Dirichlet diagrams) are discussed in more details.

Keywords: geographic information systems (GIS), distant operators and methods, Thiessen polygons, Voronoi/Dirichlet diagrams, index of spatial efficiency, spatial statistics, density of population in Slovakia

Úvod

Geografické informačné systémy (GIS) poskytujú rôznym odborom pracovné nástroje na tvorbu a analýzu priestorových modelov, ktoré sú založené na geografických konceptoch, akými sú napr. disperzia, hustota, vzdialenosť apod. Analýzy geografického priestoru, ktoré dominujú všetkým vedným odborom s prívlastkom *geo*, sa stávajú bežnou súčasťou programov GIS vo forme matematicko-štatistických a kartografických techník, resp. operátorov.

Cieľom príspevku je oboznámiť najmä so vzdialenostnými geomodelačnými technikami cez prizmu ich použitia v geografickom výskume, konkrétne v procese modelovania a analýzy priestorovej diferenciácie sídelnej štruktúry Slovenska na báze demografických charakteristík.

Priestorové nástroje geografických informačných systémov

Priestorovým analýzam a modelačných technikám v programoch GIS a ich aplikácii v geografii, resp. v humánno-geografickom výskume a praxi sa venujú práce (Kusendová 2003, Kusendová 2003a, Kusendová a Štepitová 2001).

Podľa J. Star a J. Estes (in Tuček 1996) sa priestorové analýzy, implementované do prostredia GIS, delia na dve základné skupiny na základe typu realizovaných operácií. Tie sú založené na:

a) spôsobe prepojenia medzi lokalitami identifikujúce/zohľadňujúce blízkosť, resp. vzdialenosť analyzovaných objektov (teda vytváranie rôznych priestorových zón a sieťové analýzy),
b) charakteristikách susedstva, kde pri vektorovom type usporiadania priestorových dát ide o analýzy viazané na vlastnosti susediacich plôch, resp. polygónov a pri rastrovom type usporiadania dát na susedné bunky (rôzne filtračné postupy apod.).

J. Kolář (1997) rozlišuje v integrovanej analýze priestorových dát (t. j. v spojení s tzv. atribútovými, resp. nepriestorovými dátami) tieto techniky: výberové, klasifikačné a meracie, prekrývania (angl. overlay), „neighbourhood“ operácie (prehľadávanie susedných oblastí, Thiessenove polygóny, topologické funkcie, interpolácie, generovanie izočiar) a konektívne

(meranie spojitosti, vzdialenostné analýzy, analýza sietí, funkcie šírenia, analýza viditeľnosti, osvetľovanie povrchov, priestorové zobrazovanie, resp. vizualizácia).

V programe IDRISI (Eastman 1995, http://www.idrisi.com/) je v prvej úrovni funkčného a logického rozdelenia programových modulov GIS vyčlenená skupina geografických analýz, druhá úroveň delenia modulov je podľa realizovaných operácií, kde v rámci geografických analýz je zaradená aj podskupina vzdialenostných operátorov s modulmi na výpočet vzdialeností, disperzie, sieťové a alokačné analýzy a moduly na tvorbu cenových povrchov a Thiessenových polygónov.

Vzdialenostné operátory a ich použitie

Z načrtnutých klasifikácií vyplýva, že techniky založené na vzdialenostných charakteristikách geografických objektov tvoria významnú súčasť geoinformačných analytických a manipulačných techník. Dá sa v nich rozlíšiť širší a užší výklad termínu vzdialenostné analýzy. V širšom význame tento termín označuje priestorové analýzy zamerané a využívajúce vzdialenostné charakteristiky analyzovaných objektov v danom geografickom priestore. Užšie sa týmto termínom v geoinformatike označujú priestorové analýzy vzdialenostných mier priestorových dát. Vo vektorových priestorových dátových štruktúrach ide v podstate o sieťové analýzy, ktoré sa v tomto type štruktúr sústreďujú na prvky siete v zmysle teórie grafov (uzly, hrany). Sieťové analýzy pre vektorové dáta sú súčasťou funkcií GIS, ktoré sú založené na topologických charakteristikách geografických objektov, teda na charakteristikách spojitosti (connectivity), blízkosti (proximity) alebo susedstva (neighbourhood) priestorových dát. Sieťové analýzy neumožňujú modelovanie vplyvu okolia siete na rozdiel od vzdialenostných analýz realizovaných v rastrových dátových štruktúrach, ktoré sa pre túto vlastnosť široko využívajú pri modelovaní procesov a javov spojito sa vyskytujúcich v geografickom priestore (šírenie signálu, emisií, ...). Podrobnejšie k problematike priestorových dátových štruktúr pojednáva (Pravda a Kusendová 2004) a vzdialenostných analýz v GIS (Tuček a Pacola 1996, Kusendová a Szabová 1998).

V humánnej geografii a demogeografii sa uplatňujú najmä typy analýz využívajúce vzdialenostné charakteristiky pri *štruktúrnych analýzach a modelovaní geometrických foriem* diskrétnych geografických objektov a ich *priestorových usporiadaní* (vzorov/tvarov - angl. patterns), *tvorbe a analýze* trojrozmerných *funkcionálnych* (*geo*)*štatistických polí* alebo *povrchov* (surfaces) a tvorbe *lokačno-alokačných modelov*. Pri analýze *priestorových vzorov* sa využívajú techniky GIS, ktoré globálne a lokálne rozpoznávajú, popisujú a zobrazujú (vizualizujú/animujú) analyzované dáta. Tvorba a analýza *geoštatistických polí* sa realizuje pomocou rôznych interpolačných a aproximačných techník.

Všeobecne sú nástroje GIS na priestorové spracovanie a analýzu geografických dát, v ktorých sa kladie dôraz na ich geometrické a topologické vlastnosti, v praxi najviac používané. Vzdialenostné analýzy v súčasných programoch GIS varírujú od jednoduchých štatistík geometrie geografických objektov (šírka, dĺžka, plocha, hustota, ...) cez tvorbu priestorových modelov (Thiessenove polygóny, buffering), priestorovú štatistiku v tvare rôznych testov (kvadratický, autokorelačný, ...) a indexov (najbližšieho suseda, ...) až k detekcii „povrchov" pomocou algoritmov umelej inteligencie (Kusendová 2000).

V demogeografických, ako aj ďalších aplikáciách sa tradične používajú najmä globálne extrapolačné techniky umožňujúce identifikovať všeobecné tendencie a trendy analyzovaných dát, ktoré popisujú v priestore diskrétne objekty a javy. V ostatnom čase nadobúdajú na dôležitosti lokálne interpolačné metódy identifikujúce odchýlky od všeobecných tendencií analyzovaných objektov, čo súvisí s najnovším smerom vývoja humánno-geografických výskumov (Kusendová 2003a).

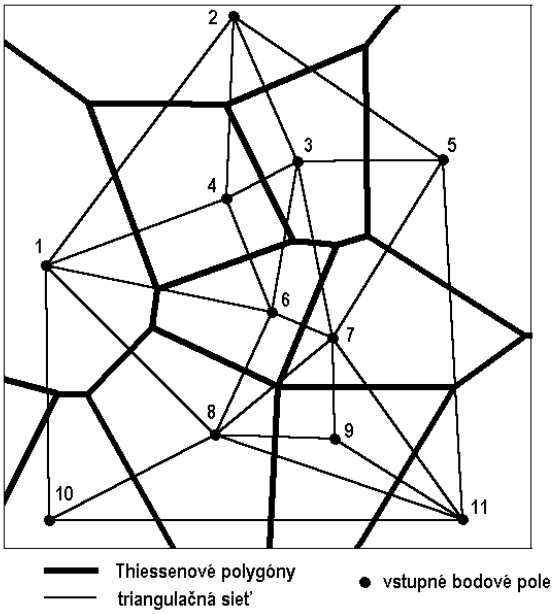
V príspevku sú vzdialenostné metódy GIS použité v procese modelovania a analýzy sídelnej štruktúry a priestorovej diferenciácii populácie Slovenska v prostredí GIS. Konkrétne

šlo o konštrukciu Thiessenových polygónov s cieľom ohraničiť spádové oblasti, resp. identifikovať nodálne regióny v sídelnej štruktúre Slovenska a vizualizovať ich populačný povrch.

Thiessenové polygóny

Thiessenové polygóny, označované podľa svojich pôvodcov aj ako *Dirichletové* alebo *Voronoi* diagramy (Sárkőzy 1998), prezentujú geoinformačnú interpolačnú a alokačnú metódu priraďujúcu plochu k vstupným bodom na základe ich vzdialenostných a smerových charakteristík. Dá sa aplikovať vo vektorových aj rastrových priestorových dátových štruktúrach. Ide o geometrickú metódu, ktorou sa rozdelí analyzovaný priestor prezentovaný bodovým dátovým poľom (povrchom) na sieť n-uholníkov (polygónov). Tie ohraničia zázemie (spádovú oblasť) každého bodu v závislosti od smeru a vzdialenosti k najbližším bodom svojho okolia. Geometrické pravidlá tvorby Thiessenových polygónov sú jednoduché: každý polygón môže obsahovať len jeden bod a všetky ostatné body, ktoré neležia v ňom, ale sú k nemu najbližšie, ležia v polygónoch tvoriace jeho hranice. Postup sa realizuje v troch krokoch:

1. Hľadajú sa najbližšie body od sledovaného vstupného bodu na základe podmienky počtu alebo vzdialenosti.
2. Každý z nájdených bodov sa spojí priamkou so sledovaným bodom, pričom sa vytvorí nepravidelná trojuholníková sieť označovaná aj ako triangulačná sieť Delona.
3. Na priamkach sa určia deliace body, ktoré sa spoja do konvexného polygónu ohraničujúci „zázemie“ sledovaného bodu (obr. 1).



Obr. 1 Thiessenové polygóny a Delonova triangulačná sieť

V rámci techník, používaných v programoch GIS, sú operácie založené na konštrukcii Voronoi diagramov štandardnou súčasťou interpolačných techník zameraných na prvotné spracovanie dát, napr. ak treba vytvoriť pravidelné vstupné bodové pole hodnôt z nepravidelného, zjednotiť rôznu polohu vstupných dát, sietí apod. Je súčasťou metódy označovanej ako *Delonova triangulácia*, ktorá tvorí základ pre tvorbu unikátnych nepravidelných trojuholníkových sietí (tzv. TIN - triangular interpolation networks). Ide o duálnu sieť k Voronoi diagramom (tenké čiary na obr. 1).

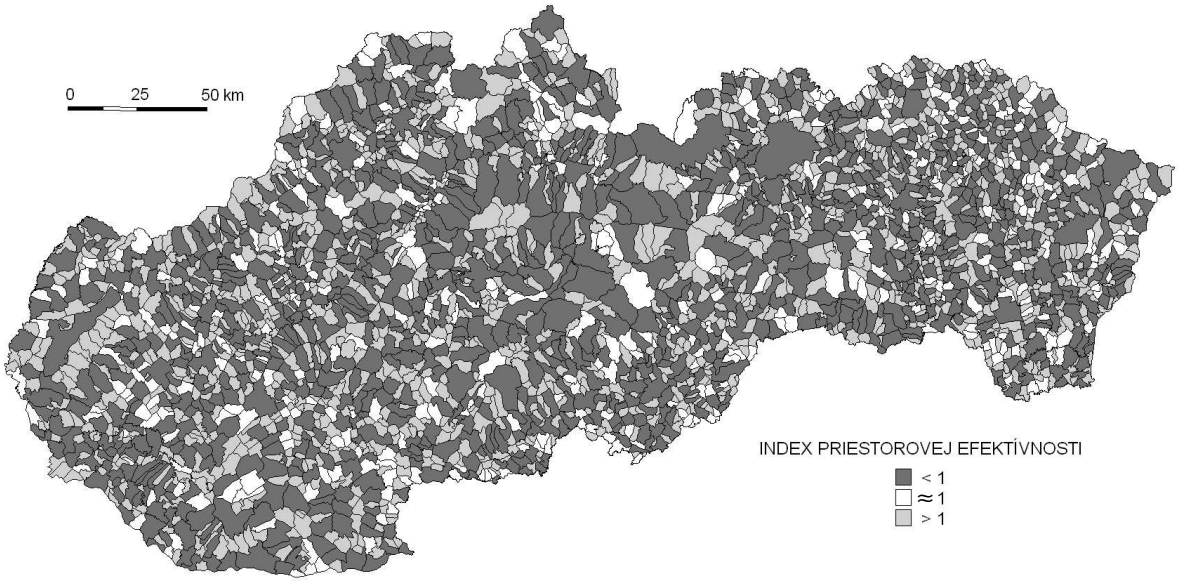
V klasifikácii často používaných interpolačných metód GIS (podľa Sárkőzy 1998) sa interpolačné metódy delia do štyroch základných skupín:

1. metódy geometrickej blízkosti,
2. štatistické metódy založené na vážených priemeroch (metóda najbližšieho suseda, inverzných vzdialeností, kriging),
3. metódy používajúce bázové funkcie (trendové, "splajnové", polynomické),
4. metódy umelých neurónových sietí) je konštrukcia Thiessenových polygónov priradená k prvej skupine (podobnejšie informácie pozri *http://www.voronoi.com.*).

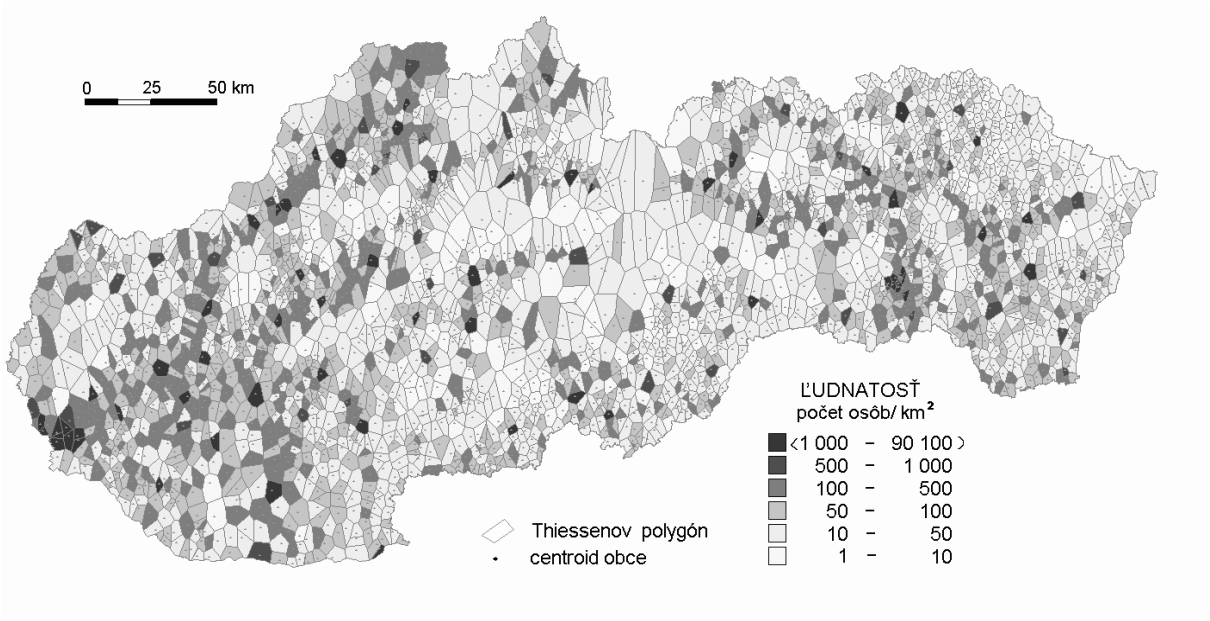
Geometria Voronoia našla svoje uplatnenie v rôznych oblastiach počínajúc kryštalografiou a končiac počítačovou kartografiou a jej vizualizačnými postupmi. V geografii použil tieto geometrické pravidlá Thiessen na vyhraničenie homogénnych klimatických regiónov na základe hydrometeorologických hodnôt získaných z bodovo lokalizovaných meteorologických staníc (odtiaľ názov Thiessenove polygóny). V humánno-geografickej a demogeografickej praxi sa využívajú na vyhraničovanie zázemí miest, spádových oblastí poľnohospodárskych, priemyselných, obchodných, rekreačných a ďalších centier. V spojení s lokačno-alokačnými postupmi vytvárajú vhodnú bázu na identifikáciu pravdepodobných oblastí vplyvu, resp. potenciálu pre rôzne aktivity človeka a modelovanie sociálno-ekonomických interakcií v geografickom priestore. Metóda má, ako väčšina geometrických metód svoje obmedzenia, napr. že rozdelenie priestoru na polygóny plne závisí od polohy vstupných bodov a nie od podstaty javu, ktorý prezentujú.

Použitie metódy Thiessenových polygónov ilustrujú obr. 2 a 3a. Bola použitá pri hraničení etalónových (ideálnych) zázemí obcí Slovenska s využitím programu MapInfo a Vertical Mapper. Získané etalóny boli ďalej analyzované a porovnávané so skutočnými hranicami, a to vizuálne, ako aj prostredníctvom *indexu priestorovej efektívnosti*, ktorý udáva veľkosť odchýlky reálnej výmery územia okresu od etalónovej výmery Thiessenovho polygónu. Cieľom komparácií bolo získať informáciu o priestorovej diferenciácii reálnych priestorových jednotiek a identifikovať priestorovo neefektívne zázemia okresov a obcí pre ďalšie analýzy.

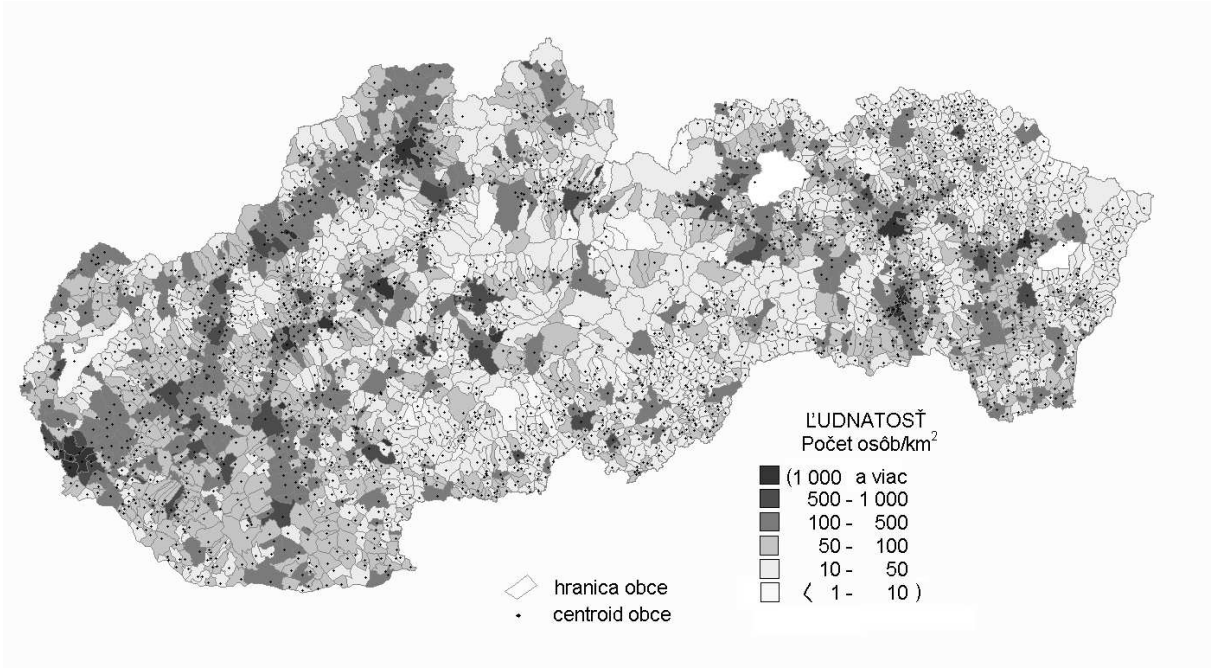
Thiesenové polygóny je vhodné použiť v kartografických prezentáciách (napr. ľudnatosti obyvateľstva) prostredníctvom kvantitatívnych areálových vyjadrovacích metód (označovaných aj ako kartogramové, resp. choropletové - Pravda a Kusendová 2004), najmä ak je značná diferenciácia veľkostí plôch geografických (fyzických) areálov (obr. 3b). Areály v príslušnom mierkovom zmenšení a transformácii zobrazujú reálne veľkosti a tvary územných jednotiek.



Obr. 2 Index priestorovej efektívnosti reálnych a ideálnych výmer obcí Slovenska



Obr. 3a Mapa ľudnatosti obcí Slovenska s použitím Thiessenových polygónov



Obr.3b Mapa ľudnatosti obyvateľstva obcí Slovenska s geografickými hranicami obcí

Záver

Použitie vzdialenostných techník a nástrojov GIS obmedzuje v súčasnosti najmä nedostatok vhodne štruktúrovaných vstupných dát. Pri výskume demogeografických, ale aj iných geografických javov pomocou priestorovej interakcie a pri riešení niektorých lokačno-alokačných problémov sú tieto techniky významnými pomocníkmi. Podmienkou ich širšieho použitia je prechod od jednoduchých foriem ku komplexnejším, heuristickým. Jednoduché techniky a operátori berú do úvahy zväčša len fyzické vzdialenostné charakteristiky. Komplexnejšie formy lokačno-alokačných modelov vyžadujú ďalšie relevantné ekonomické, sociálne a ďalšie ukazovatele, akými sú napr. reálny prepravný čas a náklady, skutočný dopyt a ponuka analyzovaných aktivít, informácie o cieľových skupinách obyvateľstva a ich pohybe, preferenciách a ďalšie. To však vyžaduje nemalé náklady na zber a odborné spracovanie primárnych dát.

Literatúra

Eastman, J. R. (1995). *IDRISI for Windows*. Worcester, USA, IDRISI Production.

Kolář, J. (1997). *Geografické informační systémy*. Praha, ČVÚT.

Kusendová D., Štepitová, D. (2001). Analýza trhu bankomatov v centre Bratislavy v prostredí geoinformačných technológií. *BIATEC*, roč. 9, 11/2001, s. 24-27.

Kusendová, D. (2000). Využitie geografických informačných systémov v humánno-geografickom výskume. In: *GIS Ostrava 2000*. Sborník referátů. Ostrava, Technická univerzita Ostrava, 236-245.

Kusendová, D. (2003). Modelling of the Spatial Distribution of Population in Slovakia with the Help of Geoinformation Techniques. In: Kowalczyk A. ed.: *Geographical space at the turn of the century – theoretical and methodological challenges*. Warsaw University, Faculty of Geography and Regional Studies, Warszawa, 2003, 39-50.

Kusendová, D. (2003a). Geografické informačné systémy a humánna geografia – vybrané teoreticko–metodologické a aplikačné aspekty. *Acta Facultatis Rerum Naturalium Universitatis Comenianae*, Geographica Nr. 44, 89-140.

Kusendová, D., Szabová, M. (1998). Vzdialenostné a sieťové analýzy – analytické nástroje GIS. *Kartografické listy*, 6, 101-110.

Pravda, J., Kusendová, D. (2004). *Počítačová tvorba tematických máp*. Bratislava, Vydavateľstvo Univerzity Komenského, 264 s.

Sárkőzy, F. (1998). *GIS functions - interpolation*. Budapest, Technical University, http://www.agt.bme.hu/public_e/funcint/funcint.html

Tuček, J. (1996). *Geografické informačné systémy*. Zvolen, Technická univerzita vo Zvolene.

Tuček, J., Pacola, E. (1996). *Školenie GIS IDRISI. Analýzy vzdialeností- časť 4*. Bratislava, Ministerstvo životného prostredia Slovenskej republiky.

Autor:
Doc. RNDr. Dagmar Kusendová, CSc., Katedra humánnej geografie a demogeografie, Prírodovedecká fakulta Univerzita Komenského v Bratislave, e-mail: kusendova@fns.uniba.sk

Index priemyselnej produkcie

Iveta Príhodová

Abstract: The aim of this article is to model the Industrial production index in Slovak Republic (adjusted from influence of number of working days) using the Box – Jenkins methodology. The series from january 1998 to march 2006 includes 99 observations.

Kľúčové slová: index priemyselnej produkcie, Box – Jenkinsova metodológia

Úvod

Cieľom príspevku je analyzovať vývoj mesačných hodnôt indexu priemyselnej produkcie Slovenskej republiky Box-Jenkinsovou metodológiou. Pre analýzy využívame mesačné hodnoty indexu od januára 1998 po marec 2006, očistené o vplyv počtu pracovných dní. Bázickým obdobím indexu, ktorý je Laspeyresovým indexom fyzického objemu, je rok 2000[1]. Index vyjadruje zmenu objemu výroby a zahŕňa ťažbu nerastných surovín, priemyselnú výrobu, výrobu a rozvod elektriny, plynu a vody. Údaje sú získavané vyčerpávajúcim zisťovaním za podniky s 20 a viac zamestnancami, za vybrané podniky do 19 zamestnancov s prevažujúcou priemyselnou činnosťou a vybrané podniky s nepriemyselnou činnosťou[2].

Metodologický aparát

1 Box – Jenkinsova metodológia[3]

1.1 Autoregresný proces *p* – teho stupňa

Autoregresný proces umožňujúci modelovať lineárnu závislosť hodnôt časového radu y_t nielen na hodnote v prvom a druhom oneskorení, ale na viacerých (p) oneskorených hodnotách sa nazýva autoregresný proces p – teho stupňa. Jeho zápis je nasledovný:

[1]Hodnoty indexu v rokoch 1998 a 1999 sú prepočítané k bázickému obdobiu, teda roku 2000.
[2]Bližšie pozri www.statistics.sk
[3]Uvádzame len procesy použité vo výslednom tvare modelu. Ďalej pozri napr. BOX, G.E.P. – JENKINS, G.M. Time Series Analysis: Forecasting and Control. New Jersey: Prentice Hall International, Inc. 1994.

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + a_t \tag{1.1}$$

Tento proces možno pomocou spätného operátora zapísať ako:

$$\left(1 - \phi_1 B - ... - \phi_p B^p\right) y_t = a_t \tag{1.2}$$

Proces AR (*p*) je stacionárny, ak korene polynomickej rovnice $\left(1 - \phi_1 B - ... - \phi_p B^p\right) = 0$ ležia mimo jednotkového kruhu.

Rozptyl procesu AR(*p*) možno pomocou σ_a^2, ϕ_1, ..., ϕ_p pri známych hodnotách ρ_1,..., ρ_p vyjadriť ako:

$$\gamma_0 = \frac{\sigma_a^2}{1 - \phi_1 \rho_1 - ... - \phi_p \rho_p} \tag{1.3}$$

Hodnoty autokorelačnej funkcie získame zo vzťahu:

$$\rho_k = \phi_1 \rho_{k-1} + ... + \phi_p \rho_{k-p}, \text{ pre } k = 1,2,... \tag{1.4}$$

Ak sú korene rovnice $\phi_p(B) = 0$ reálne, ACF exponenciálne klesá k nule. V prípade, ak sú korene uvedenej rovnice komplexné, ACF exponenciálne klesá k nule v sínusových vlnách.

Hodnoty PACF sú pre oneskorenie *k = 1,2,...,p* rôzne od nuly, pre ďalšie oneskorenia sa nule rovnajú.

Procesy kĺzavých priemerov vyššieho ako druhého stupňa sa často vyskytujú pri modelovaní časových radov finančných ukazovateľov, napr. cien akcií, akciových indexov a menových kurzov. V týchto modeloch umožňujú často zachytiť výkyvy spojené s rôznou intenzitou obchodovania v jednotlivých dňoch týždňa.

1.2 Sezónny autoregresný proces *P*-teho stupňa SAR(*P*)$_s$

Pomocou predchádzajúcich tvrdení môžeme SAR(*P*)$_s$ proces zapísať nasledovne:

$$y_t = \Phi_1 y_{t-s} + \Phi_2 y_{t-2s} + ... + \Phi_P y_{t-Ps} + a_t \tag{1.5}$$

Vychádzame z predpokladu, že a_t nezávisí od hodnôt y_{t-1}, y_{t-2},...

Spätný operátor umožňuje nasledujúci zápis procesu:

$$\left(1-\Phi_1 B^S-\Phi_2 B^{2S}-...-\Phi_P B^{PS}\right)y_t = a_t\,, \tag{1.6}$$

resp.

$$\Phi_P\left(B^S\right)y_t = a_t \tag{1.7}$$

Proces je konvergentný, ak korene rovnice $\Phi_P\left(B^S\right)=0$, ležia vo vnútri jednotkového kruhu.

Tento proces je stacionárny, ak ho možno prepísať do tvaru sezónneho lineárneho procesu:

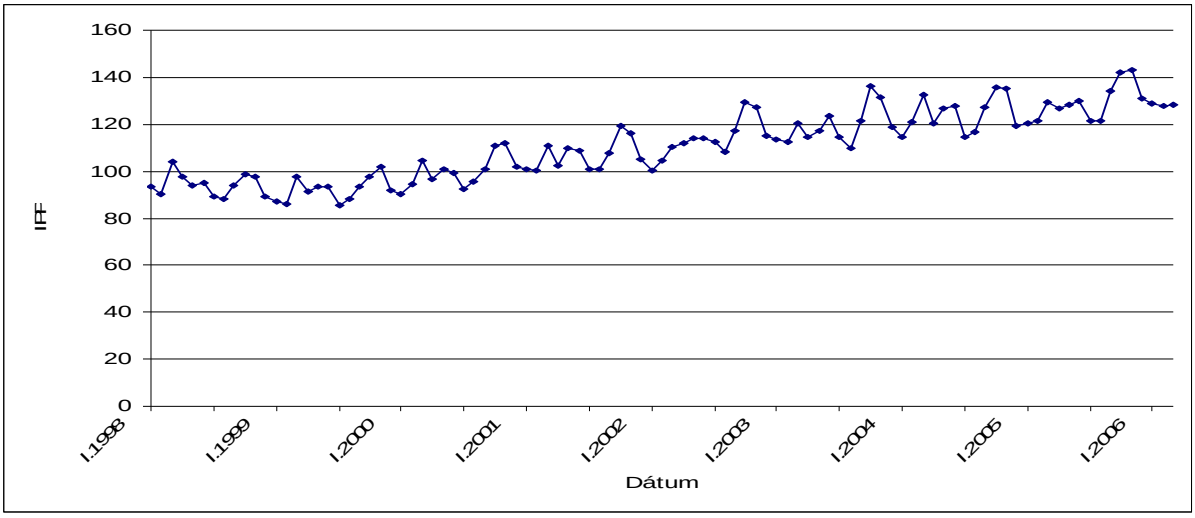
$$y_t = \Psi\left(B^S\right)a_t \tag{1.8}$$

Priebeh autokorelačnej i parciálnej autokorelačnej funkcie je pre *P > 1* zložitý a jeho systematickosť je pomerne ťažké opísať.

2 Modelovanie vývoja indexu priemyselnej produkcie

Časový rad indexu priemyselnej produkcie nie je stacionárny v strednej hodnote (Graf 1). Hypotézu o potrebe stacionarizovania časového radu potvrdzuje výstup upraveného Dickey – Fuller (Augmented Dickey – Fuller test) testu jednotkového koreňa (Tabuľka 1). Na každej zo štandardne testovaných hladín významnosti prijímame hypotézu o prítomnosti jednotkového koreňa v tomto časovom rade. Na jeho stacionarizovanie, ktoré je predpokladom následnej aplikácie Box – Jenkinsovej metodológie, použijeme diferencovanie. Časový rad vykazuje okrem prítomnosti trendovej zložky i prítomnosť zložky sezónnej. Na stacionarizovanie časového radu v strednej hodnote sme použili prvú sezónnu diferenciu. ADF test jednotkového koreňa diferencovaného časového radu potvrdil dostatočné stacionarizovanie časového radu.

Graf 1 Vývoj indexu priemyselnej produkcie od januára 1998 do marca 2006



Tabuľka 1 ADF test jednokového koreňa

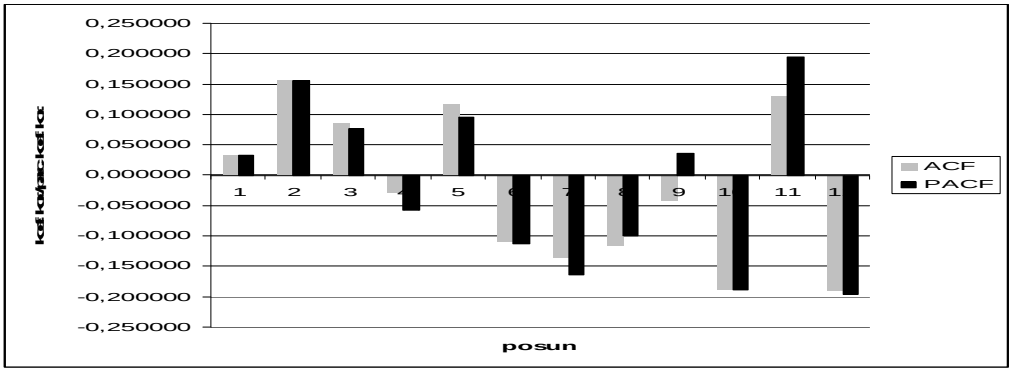
ADF Test Statistic	-2.503391	1% Critical Value*	-3.4986
		5% Critical Value	-2.8912
		10% Critical Value	-2.5824
*MacKinnon critical values for rejection of hypothesis of a unit root.			

Po stacionarizovaní časového radu v strednej hodnote sme pristúpili k modelovaniu vývoja strednej hodnoty tohto radu. Vhodným modelom vývoja je SARIMA$(1,0,0)(0,1,1)_{12}$ model s konštantou. Podmienka stacionarity autoregresného i sezónneho autoregresného procesu prvého stupňa je splnená. Hypotézu o štatistickej významnosti konštanty modelu a parametra ϕ_3 nesezónneho autoregresného procesu kĺzavých priemerov môžeme prijať na akejkoľvek hladine významnosti, rovnako ako v prípade odhadu parametra Φ_{12} sezónneho autoregresného procesu kĺzavých priemerov. Odhad parametra ϕ_1 je štatisticky významný na hladine vyššej ako 0,0215. Na základe hodnoty D – W štatistiky môžeme rezíduá modelu považovať za nezávislé. Hypotézu o nezávislosti rezíduí modelu potvrdzuje i graf autokorelačnej a parciálnej autokorelačnej funkcie rezíduí modelu (Graf 2). Žiaden z odhadnutých koeficientov autokorelácie, resp. parciálnych koeficientov autokorelácie nepresahuje hranicu $\pm 2/T^{-0,5}$. Odhadnutý model je štatisticky významný na akejkoľvek hladine významnosti. Na základe J – B testu normality prijímame i hypotézu o normalite rozdelenia rezíduí odhadnutého modelu (p – value 0,6256).

Tabuľka 2 Odhad parametrov modelu SARIMA(1,3),0,0)(0,1,1)$_{12}$

Method: Least Squares				
Sample(adjusted): 2000:01 2005:06				
Included observations: 66 after adjusting endpoints				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.751596	0.578393	9.944101	0.0000
AR(1)	0.267766	0.113490	2.359387	0.0215
AR(3)	0.298228	0.111610	2.672062	0.0096
AR(12)	-0.232246	0.081292	-2.856933	0.0058
R-squared	0.312507	Mean dependent var		6.012121
Adjusted R-squared	0.279241	S.D. dependent var		3.651154
S.E. of regression	3.099740	Akaike info criterion		5.159205
Sum squared resid	595.7201	Schwarz criterion		5.291912
Log likelihood	-166.2538	F-statistic		9.394229
Durbin-Watson stat	1.920696	Prob(F-statistic)		0.000033

Graf 2 Autokorelačná a parciálna autokorelačná funkcia rezíduí modelu



Testom Lagrangeovho multiplikátora sme overili prítomnosť podmienenej heteroskedasticity v druhých mocninách rezíduí odhadnutého modelu. Hypotézu o existencii podmienenej heteroskedasticity môžeme na základe tohto testu zamietnuť na akejkoľvek hladine významnosti. Lineárny model strednej hodnoty nie je potrebné rozšíriť o nelineárny model podmieneného rozptylu.

Záver

Cieľom príspevku bolo nájsť vhodný ARIMA model vývoja indexu priemyselnej produkcie. Zistili sme, že vhodným modelom vývoja indexu priemyselnej produkcie je model SARIMA(1,3),0,0)(0,1,1)$_{12}$ s konštantou. Rezíduá tohto modelu majú vlastnosti bieleho šumu. Prognóza indexu priemyselnej produkcie na nasledujúce obdobie (apríl 2006) je 130,8 %.

Literatúra:

ARLT, J. Moderní metody modelování ekonomických časových řad. Praha: Grada Publishing, 1999.

BOX, G.E.P. – JENKINS, G.M. Time Series Analysis: Forecasting and Control. New Jersey: Prentice Hall International, Inc. 1994.

CHAJDIAK, J. – KOMORNÍKOVÁ, M. – IŠTVÁNIKOVÁ, A. Metódy prognózovania. Bratislava: STATIS, 1998.

RUBLÍKOVÁ, E. – ARLT, J. – ARLTOVÁ, M. – LIBIČOVÁ, L. Analýza časových radov. Zbierka príkladov. Bratislava: Ekonóm, 2001.

RUBLÍKOVÁ, E. Prognostická štatistika. Adaptívne metódy. Bratislava: Edičné stredisko EU, 1989.

Kontakt:

Ing. Iveta Príhodová

Ekonomická univerzita v Bratislave

Fakulta hospodárskej informatiky, Katedra štatistiky

Dolnozemská cesta 1/B

852 35 Bratislava

e-mail: i_prihodova@yahoo.de

DATAMINING

Marianna Vavrová

Abstrakt: Datamining sits at the interface between statistics, computer science, artificial intelligence, machine learning, database management and data visualization. To date, most of the above types of information (for whatever application domain) have not been exploited. Patterns that are "interesting" normally concern relationships that are not obvious, or are unexpected. There is simply too much information to process with existing tools. New tools and techniques need to be developed. This is the motivation for the research.

1. Úvod do problematiky - Manažérske systémy

Manažérske systémy spadajú do oblasti, pre ktorú sa dnes vžilo označenie Data Mining/Data Warehousing. A práve manažéri na všetkých úrovniach riadenia sú koncoví užívatelia nástrojov a systémov DM/DW a sú to práve ich nároky na spracovanie, reporting a analýzu dát a informácií, ktoré ovplyvňujú vývoj týchto nástrojov. Ich potreby v dnešnej dobe narastajú s tým, ako sa zostruje konkurencia, zväčšujú sa objemy dát, ktoré je nutné analyzovať, zvyšujú sa nároky na rýchlosť‘ analýzy a prijatia následného rozhodnutia.

V minulosti plne dostačoval ako zdroj informácií pre manažérov tabuľkový editor s exportom vybraných dát. Pokiaľ sa pozrieme ešte ďalej do minulosti, ani papierové reporty a kreslenie grafov na tabuľu nie sú až takou dávnou minulosťou.

S tým, ako sa zväčšuje v jednotlivých segmentoch konkurencia a ako sa firmy vybavujú komunikačnými a informačnými technológiami, sa rozhodovanie posúva z horizontu dní a týždňov do oblasti hodín, alebo v kritických situáciách do rádov desiatok minút.

Napr. oddelenie predaja nie je schopné poskytnúť do hodiny údaje o predajnosti 20 druhov tovaru za posledný mesiac spolu s vyhodnotením nákladov na ich distribúciu, ak je potrebné „vyťahovať“ dáta z 2-3 databáz, dávať ich dohromady a vyhodnocovať ich v tabuľkovom procesore. Pokiaľ je však k dispozícii dátový sklad, obsahujúci dáta z produkčných systémov a nad ním manažérska nadstavba, kde sa jednoduchým zadaním parametrov zobrazia požadované prehľady, bude hodina na spracovanie stačiť.

Potreba rýchle dostupných a hlavne neustále aktuálnych informácií je vidieť vo väčšine odvetví, avšak zvlášť markantné v oblasti telekomunikácií, finančných a bankových inštitúcií, atď... Vedenia podnikov, ktoré chcú byť v súčasnom svete úspešné, sa musia rozhodovať okamžite na základe vzniknutej situácie.

Datamining

Tento pojem, slovensky prekladaný ako „dolovanie dát“, je citovaný najčastejšie v súvislosti s dátovými skladmi. Môžeme sa stretnúť s dvoma významami tohto pojmu:

● ***Širší význam*** - v širšom zmysle sú s týmto pojmom spájané všetky „zložitejšie“ aktivity nad dátovým skladom.

●***Užší význam*** - v užšom zmysle je možné dolovanie dát charakterizovať ako špecifický proces získavania dát pre rozhodovanie z veľmi rozsiahlych dátových skladov a to extrakciou

relevantných, predom neznámych informácií. Cesty, ako toto dolovanie nových informácií realizovať sú:

• Generalizácia a sumarizácia dát;
• Využívanie asociačných pravidiel;
• Klasifikácia dát;
• Zhlukovanie dát.

V tomto užšom zmysle musia byť preto k dispozícii vyspelé nástroje pripravené špecialistami, a koncový užívateľ je skôr v úlohe, kde dostáva nové a „často predom neznáme“ informácie.

Snahy získať z bohatých fondov dátových skladov nové neznáme informácie sú spájané z radov nových pojmov ako „knowledge extraction“ (extrakcia znalostí), „data archaelogy“ (archeológia dát), ale najviac sa ujal pojem „data mining“. Otvorila sa tak možnosť, aby celý rad teoretických výskumov z oblasti matematickej štatistiky, strojového učenia, umelej inteligencie a neurónových sietí bola zhmotnená do nových nástrojov, ktoré sa pod hlavičkou „data mining“ dostávajú na stoly manažérov.

1.1 Informačné technológie a Datamining

Tento termín môže pripomínať každodenný boj o informácie, alebo proces získavania údajov pre podnikový systém od špecializovaných útvarov a používateľov. Druhým extrémom je predstava zázračného prostriedku, ktorý po znesení všetkých dát z podniku na jedno miesto sám nájde riešenia všetkých problémov.

- čo to vlastne Dataminig je?
- aké metódy a techniky sa využívajú?
- niektoré reálne úlohy riešené v rôznych odvetviach.

Súčasný svet je charakterizovaný explóziou objemu dát zbieraných a ukladaných do databáz. S požiadavkami na zhromažďovanie informácií sa stretávame vo všetkých oblastiach - od služieb (objednávky zásielkových služieb či cestovných kancelárií, rezervácia cestovných lístkov, leteniek), cez bankovníctvo, telekomunikácie, štátnu správu, až po zdravotníctvo a poisťovníctvo. V bankovníctve ide napr. o archiváciu bankových transakcií, žiadostí o úver, bankových splátok, v oblasti telekomunikácií zase o informácie o telefonickej prevádzke a platbách za ne (v prípade mobilných telefónov informácie o zemepisnej polohe, atď...) Štátna správa je asi najväčším „zberateľom informácií“, nech už sa jedná o daňové priznania, žiadosti o sociálnu podporu, najrôznejšie registre a celý rad iných oblastí.

V poslednej dobe začína hrať dôležitú úlohu v oblasti koncového predaja práve zber údajov o zákazníkoch a to najviac v supermarketoch. Najrôznejšie zákaznícke karty slúžia nielen pre získanie množstevných zliav, ale hlavne pre vytvorenie databanky informácií o nákupe jednotlivých druhov tovaru. V oblasti poisťovníctva ide o registráciu poistiek, v zdravotníctve zase o archiváciu zdravotných záznamov alebo údajov pre zdravotné poisťovne.

Ak u nás ešte nie sú niektoré vyššie popisované zdroje dát, bežné v elektronickej podobe, tak vo vyspelých štátoch, pokiaľ v elektronickej podobe priamo nevznikajú, tak v nej určite

končia, uložené v databázach. Táto záplava dát je obhospodarovaná transakčnými systémami, ktoré väčšinou spracovávajú aktuálne transakcie, poprípade je postúpená do systémov navrhnutých pre analýzy (systémy pre dodávanie informácií - Information Delivery), ktorých úlohou je poskytovať prehľadné informácie pre rozhodovanie. Základom moderných systémov na dodávanie informácií je dátový sklad (Data Warehouse) - centrálne skladisko zjednocujúce zber informácií z celého podniku a ukladanie historický dát, to všetko optimalizované pre rôzne analýzy.

Spracovanie dát z rozsiahlych databáz a dátových skladov má v dnešnom svete informačných technológií najrôznejšie formy. Tradičné prístupy analyzujúce dáta prostredníctvom zostáv a výkazov sú dnes väčšinou založené na dotazovacích nástrojoch (SQL), pracujúcich nad relačnými databázami, prípadne na technikách označovaných ako OLAP (On-Line Analytical Processing), ktoré často využívajú uloženie dát v multidimenzionálnych databázach k rýchlej prezentácii dát vo forme tabuliek sumovaných pre rôzne hierarchicky usporiadané dimenzie (rok, štvrťrok, mesiac verzus kraje, okresy alebo obce).

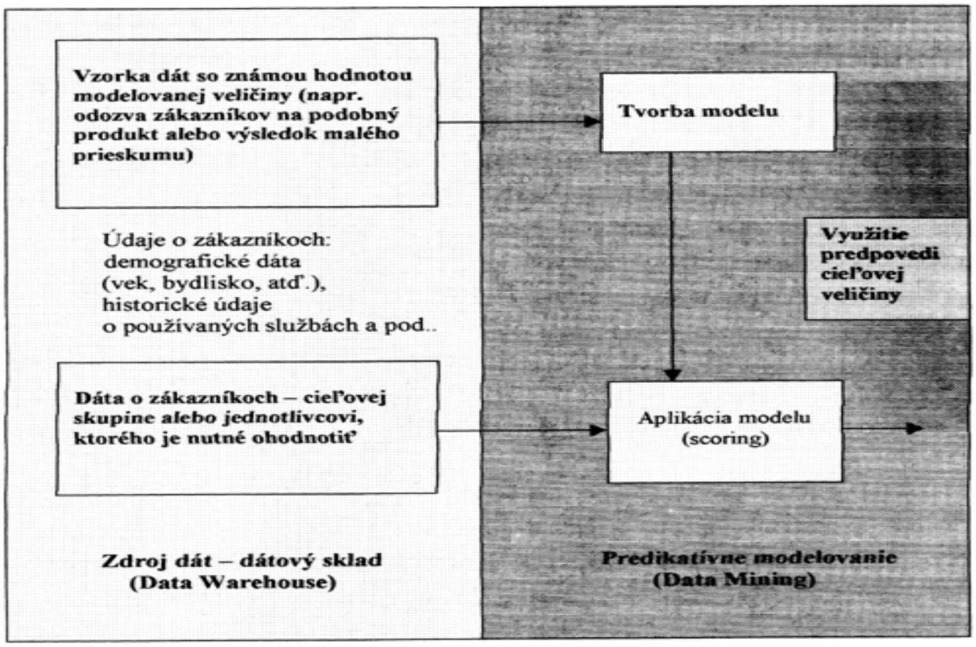
Tieto techniky umožňujú udržiavať prehľad o okamžitej pozícii podniku, či rýchlu prípravu konsolidovaných finančných riportov - to všetko aj v rozsiahlych organizáciách a za čas, za ktorý by sa pred niekoľkými rokmi zdala nesplniteľná. Napriek tomu je veľa úloh, na ktoré tieto bežné postupy nestačia. Vo väčšine týchto úloh nie je špecifikovaná konkrétna, možno aj zložitá otázka na obsah databázy. Naopak, cieľom je nájsť, ktoré údaje, či skôr ich kombinácie, sú dôležité pri komerčnom využití dát z databázy. Pokiaľ je obsahom tradičnej databázovej otázky otázka „V ktorom kraji bol v tomto štvrťroku najúspešnejší predajca tovaru typu X?“, tak sa úlohy dolovania dát snažia nájsť riešenia problémov - „Aké podmienky (skladba produktov, demografie cieľovej skupiny) zabezpečuje dlhodobo najlepšie výsledky predaja?“.

1.2 Definícia dolovania dát

Všeobecná definícia dolovania dát je popisovaná ako proces výberu, prehľadávania a modelovania vo veľkých objemoch dát, slúžiacich k odhaleniu skôr neznámych vzťahov medzi dátami, za účelom získavania obchodnej výhody. Obchodná výhoda je teda označovaná ako cieľ dolovania dát - dolovanie by vždy malo mať za cieľ riešenie konkrétneho obchodného problému či nájdenia cesty k vylepšeniu procesu. Cieľ by mal byť dopredu definovaný a iba na jeho základe by sa mali pripravovať dáta.

Pojem veľký objem dát môže samozrejme znamenať‘ rôzne objemy v rôzne veľkých organizáciách, ale zdôrazňuje nutnosť‘ podpory managmentu a informačných technológií - IT pri sprístupňovaní rozsiahlych dát z rôznych častí organizácie. Dáta pre dolovanie by mali byť brané z dátového skladu zahrňujúceho historické hodnoty z rôznych podnikových systémov. Nasadenie technológie dolovania dát predpokladá prípravu podnikových procesov umožňujúcich kontinuálne využívanie analýz a podporujúcich spätné väzby od používateľov. Tie potom môžu ovplyvňovať ako proces zberu dát, tak definíciu nových cieľov. Vedúci firmy, dodávajúcej riešenia zahrňujúce dolovanie dát, majú pre tento účel spracované metodológie procesu dolovania, ktoré pomáhajú, ako s vlastnými projektami, tak s riadením nadväzujúcich podnikových procesov. (proces dolovania dát znázorňuje nasledovná schéma 1.)

Schéma 1. Znázornenie procesu dolovania dát



2. Dolovanie dát a dátové sklady

Existuje veľa dôvodov pre úzku naviazanosť dátového skladu a dolovania dát. Asi najdôležitejším z nich je kvalita vstupných dát pre dolovanie. Sebadokonalejšia modelovacia technika či analýza neprinesú očakávaný výsledok, pokiaľ nie sú vstupné dáta očistené od chýb, skontrolovaná úplnosť‘ všetkých požadovaných údajov a zjednotené formáty z rôznych systémov. Procesný charakter dolovania dát vyžaduje, aby sa ako vstup dolovania dát používali priebežne aktualizované dáta. Všetkým týmto požiadavkám vyhovuje dátový sklad. Rozšírenie dátových skladov je naopak jedným z hybných prvkov búrlivého rozvoja technológie dolovania dát.

2.1 Metódy dolovania dát

Dolovanie dát je označením triedy úloh, ktoré riešia niekedy zdanlivo nesúvisiace problémy z najrôznejších oborov. Je pozoruhodné, koľko praktických aplikácií má niekoľko obecných metód analýzy dát. Výber metódy, ktorá bude použitá pre riešenie daného problému, je jedným z krokov procesu dolovania dát. Na zreteli je potrebné mať cieľ, pre ktorého splnenie môžeme použiť viac metód. Potom je dobré poznať ich výhody a mať možnosť porovnať ich výsledky.

Prediktívne modelovanie

Je postup, kedy sa na základe známej množiny vstupných a známych im odpovedajúcich výstupných hodnôt, hľadá najpravdepodobnejšia hodnota výstupu pre predom neznáme kombinácie vstupných hodnôt. Elementárnym príkladom prediktívneho modelovania je napr. hodnotenie rizika úveru v bankovníctve. Banka sústreďuje veľa záznamov o svojich minulých

klientoch a samozrejme vie, ktorí sú zlými dlžníkmi. Po vytvorení prediktívneho modelu popisujúceho hodnotenie dlžníka (výstup) na základe informácií o ňom (vstupné dáta), je možné ohodnocovať riziká nových, prichádzajúcich zákazníkov (o ktorých je možné zistiť údaje používané ako vstupné dáta modelu). Používanými technikami pre prediktívne modelovanie sú rôzne typy regresie, neurónové siete a rozhodovacie stromy.

Regresia

Je štandardná štatistická metóda schopná popísať stupeň dôležitosti vstupných premenných na výstup. Jej sila tkvie v teoretickej prepracovanosti odhadu chýb modelu a možnosti hľadať i závislosť na kombinácii vstupných premenných. Dobre sú prepracované metódy regresie pre odhad binárneho výstupu (logistická regresia) či regresia pracujúca s dátami nadobúdajúcimi len diskrétne (aj nečíselné) hodnoty. Použitie regresie je limitované prácnosťou a časovou náročnosťou vývoja zložitejších modelov.

Neurónové siete

Predstavujú novú modernú techniku prediktívneho modelovania, vynikajúcou veľkou variabilitou možných modelov a jednoduchosti modifikácie ich návrhu. Z pohľadu regresie sú neurónové siete elegantnou technikou pre hľadanie parametrov modelu, založeného na veľmi flexibilnom systéme vnorených funkcií. Na druhej strane model založený na neurónovej sieti nemá zrozumiteľnú interpretáciu.

Rozhodovacie stromy

Získali popularitu vďaka svojej jednoduchej interpretácii. Popis modelu pomocou rozhodovacieho stromu je radom jednoduchých rozhodovacích pravidiel, často prezentovaných vo forme grafu. Tieto grafy môžu byť ľahko bez hĺbkových znalostí štatistických metód interpretované riadiacimi pracovníkmi. Pri použití všetkých techník modelovania je nutné nešiť problémy s voľbou počtu parametrov modelu, ich konvergencie a odhadu chýb.

Klasifikácia

Všeobecne je klasifikácia metódou pre rozdeľovanie dát do skupín podľa istých kritérií. Pokiaľ sú tieto kritéria dopredu známe, aspoň pre vzorku dát, je možné pomocou metód prediktívneho modelovania vyvinúť model, ktorého výstupom je klasifikačná premenná. Oveľa častejším prípadom je neriadená klasifikácia, kedy výsledné kritériá nie sú predom známe a úlohou klasifikácie je ich nájdenie. Používanou technikou v takýchto prípadoch je zhluková analýza (Cluster Analysis). Elementárnym príkladom zhlukovej analýzy je napr. nájdenie skupín obchodov na základe ich obratu, sortimentu a typu zákazníkov. Nájdené skupiny môžeme potom využiť napr. pre špecifikáciu reklamnej kampane, zameranej na jednotlivé skupiny predajní.

Analýza asociácii

Najčastejším použitím analýzy asociácii, a zároveň jej ilustračným príkladom, je tzv. analýza nákupného košíka. Tá sa zaoberá hľadaním kombinácií produktov, ktoré sa vo

vstupných dátach (nákupnom koši spotrebiteľov) vyskytujú významne častejšie spolu. Cieľom je odhaliť pravidlá typu: pri nákupe tovaru A a C spotrebitelia výrazne častejšie nakupujú tovar D a B. Odhalenie takýchto kombinácií pomáha marketingovým odborníkom v organizovaní ponuky či spoločných balíčkov produktov.

Vzorkovanie

Vzorkovanie nie je algoritmom riešiacim priamo nejakú zadanú úlohu dolovania dát, ale je to jedna zo základných techník dolovania dát, umožňujúcich získať‘ výsledok v rozumnom čase. Vzorkovanie je výber obmedzenej množiny dát zo základného súboru. Najjednoduchším spôsobom vzorkovania je náhodný výber, ktorý slúži iba ku zmenšeniu objemu spracovávaných dát a tým k zrýchleniu výpočtov. Zložitejšie metódy vzorkovania, napr. výber rovnakého počtu záznamov daného typu (rovnaký počet záznamov z každého tarifného pásma), umožňujú redukciu objemu spracovávaných dát pri súčasnej záruke požadovanej presnosti výsledku. Vzorkovanie databázy pre analýzy je iste neobvyklé z pohľadu klasických databázových aplikácií užívajúcich SQL nástrojov, ale objavujú sa tiež úvahy o nutnosti zaradenia takýchto služieb do štandardných databázových serverov.

2.2 Použitie techník dolovania dát

Treba si uvedomiť, že niektoré oblasti aplikácie dolovania dát zostávajú dôverným firemným tajomstvom. To isté platí i pre väčšinu získaných výsledkov. Popis modelového prípadu je v nasledujúcom texte uvedený na analýze úverového rizika.

Analýza úverového rizika - výber a overovanie kandidátov žiadajúcich o úver, môžeme opäť popísať prediktívnym modelom, založeným na známom správaní sa nastávajúcich klientov. Výhodou v tomto prípade je znalosť množstva dát o klientoch.

• Vyhodnocovanie marketingových kampaní - tvorbu prediktívneho modelu odozvy, získaného na základe zo vzoriek zákazníkov, môžeme vykonať výber z rozsiahlej databázy zákazníkov, ktorý garantuje s najväčšou pravdepodobnosťou odozvy.

• Analýza odchodu zákazníkov (churn) - prediktívny model získaný analýzou dát o zákazníkoch môžeme použiť‘ pre plánovanie akcií, ktoré môžu zabrániť odchodu nastávajúcich, najrizikovejších zákazníkov. V telekomunikáciách je používaný pojem „churn“ pre zmenu poskytovateľa služieb.

• Segmentácia zákazníkov - rozdelenie zákazníkov do skupín pre marketingové účely. Segmenty potom môžu definovať rôzne cieľové skupiny.

• Detekcia podvodov - pomocou prediktívneho modelovania (najčastejšie neurónové siete), či zhlukové analýzy, môžeme odhaliť podozrivé správanie sa pri platobnom styku.

• Analýza produktov - príma aplikácie analýzy asociácií - umožňuje definovať komplementárne produkty pre dané segmenty zákazníkov. Potom je možné cielene oslovovať zákazníkov, ktorým chýba časť portfólia produktov, či zostavovať požadované balíčky služieb.

• Analýza správania sa zákazníkov - predikcia napr. vývoja dopytu na základe historických dát.

• Analýza sekvencií - výber najčastejšie sa vyskytujúcich postupností, či hľadania stavov predchádzajúcich nejaké udalosti (poškodenie iniciované viacerými vplyvmi).

3. Štatistika a Datamining

Data Mining (DM) spája výpočtovú vedu a štatistiku zhrňujúc výhody oboch disciplín na vytvorenie postupu extrakcie informácií z veľkých databáz. Je to oblasť, ktorá v poslednej dobe priťahuje pozornosť. Tento článok sa zameriava na štatistické termíny, ktoré sú priamo relevantné pre DM a pokúša sa upozorniť na spojenia medzi štatistikou a výpočtovou vedou poskytujúc možnosti pre budúci postup v analýze dát.

Štatistika sa zaoberá zberom, organizáciu a vyhodnocovaním číselných dát, ale najmä analyzovaním charakteristík populácie odvodzovaním z jednotlivých vzoriek. Oblasť DM sa týka „učenia z dát“ alebo „vyťahovania informácií z databáz“. Kontext obsahuje štatistiku, ale tak trochu s iným náhľadom. V podrobnostiach DM zahŕňa spätnú analýzu dát: tak námet, ako experimentálny náčrt sú vedľajšou oblasťou DM a spadajú pod štatistickú. Dobyvateľov dát viac zaujíma porozumenie než presnosť alebo predvídavosť. Ohniskom toho je jednoduchá interpretácia modelov zahrňujúcich pravidlá, stromy, grafy.... V aplikáciách DM sa pracuje s množstvom premenných obrovských rozmerov, čo vyžaduje efektívnosť výpočtov, pričom výstup môže byť sekundárnou úlohou. Zo skúseností v DM je často uprednostňované zameriavanie sa na vzorky (pattern-focused) pred budovaním modelu (model-focused), napr. radšej ako vybudovať globálny model so všetkými prepojeniami medzi premennými, DM algoritmu bude produkovať nastavenie prepojení iba medzi lokálnymi premennými. Súčasný DM je veľmi poháňaný záujmom o výpočtovú techniku. Avšak zameraním sa výhradne na výpočtové problémy je jednoduché zabudnúť na štatistiku, ktorá je jadrom problému. Termín DM mal dlho vedľajší význam v štatistickej literatúre. Cieľom článku je presvedčiť čitateľa, že moderná štatistika ponúka významné rady pre DM, hoci množstvo problémov zostáva nevyriešených.

4. Firmy zaoberajúce sa problematikou Datamining

SAS Institute - spoločnosť má vyše 20-ročné skúsenosti v oblasti vývoja softwaru na dodávanie informácií (Information Delivery). Firma ponúka ucelený rad nástrojov pre tvorbu a riadenie dátových skladov, analýzy, modelovanie a prezentáciu dát. Produkty podporujú prostredia AIX, HP-UX, Solaris, Windows 95/NT. Enterprise Miner umožňuje bez hlbokých teoretických znalostí vytvárať prediktívne modely (k dispozícii sú regresie, neurónové siete, a rozhodovacie stromy), analýzu zhlukov, asociáciu i sekvenciu. Je ľahko integrovateľný s dátovým skladom, ponúka flexibilné sady rôznych techník a grafické výstupy zamerané na obchodných používateľov. Skúsenosti SASu s realizáciou projektov na dolovanie dát sú zhrnuté do metodológie Data Mining Project Methodology.

Silicon Graphics - SGI Mineset sa vyznačuje silnými vizualizačnými nástrojmi pre komplexné dátové sady. Ponúka pokročilú 4D vizualizáciu - 3D grafy sa môžu meniť podľa 2-och ďalších premenných dimenzií. Podporuje prácu v prostrediach Unix a Windows NT,

podporu prístupu k dátam v databázach Oracle, Informix a Sybase a jednoduchú integrovateľnosť do prostredia intranetu. Zmeny zaznamenali i analytické metódy, kde môžeme nájsť napr. zhlukovú analýzu, regresné a rozhodovacie stromy, rozhodovacie tabuľky alebo analýzy evidencie.

IBM - Intelligent Miner poskytuje rôzne techniky, ktoré sú integrované do jedného modelu. Ponúka najlepší clustering. Beží v prostredí AIX, MVS, OS/400, OS/2 a Windows NT/95.

Literatúra a použité zdroje

[1] Agrawal R., Mannila H., Srikant R., Toivonen H., Verkamo A. I.: Fast discovery of association rules. In: (Fayyad, Piatetsky-Shapiro, Smyth, Uthurusamy, eds) Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996

[2] Berka P.: Dobývaní znalostí z databází, Academia Praha, 2003

[3] Fayyad U., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. eds: Advances in Knowledge Discovery and Data Mining. AAAI Press/MIT Press, 1996

[4] Frawley W. J., Piatetsky-Shapiro G., and Matheus C. J. "Knowledge Discovery in Databases: An Overview", Knowledge Discovery in Databases, Piatetsky-Shapiro and Frawley (eds.), AAAI/MIT Press, 1991, pp. 1-27.

[5] Chajdiak J. a kol.: Výpočtová štatistika. ES VŠE, Bratislava, 1988

[6] Chajdiak J. a kol.: Štatistický analytický system. STATIS, Bratislava, 1992

[7] Chajdiak J., Rublíková E., Gudába M: Štatistické metódy v praxi. STATIS, Bratislava, 1994

[8] Kouba Z.: Datové sklady, Dobývání znalostí z databází, Sborník přednášek, FIS VŠE Praha, 2000

[9] Kantardzic M.: Data Mining – concepts, models, methods, and algorithms, IEEE Press, New York, 2003

[10] Klosgen W., Zytkow J.: Knowledge Discovery and Data Mining. Tutorial Notes. PKDD´97, Trondheim 1997

http://www.wizsoft.com
http://www.sybase.com. www. sybase. sk
http://www.oracle.com
http://www.datamining.com
http://www.dw-institute.com

Kontakt:

Ing. Marianna Vavrová, externá doktorandka
Katedra štatistiky - Fakulta hospodárskej informatiky
Ekonomická univerzita v Bratislave
Dolnozemská cesta 1
852 35 Bratislava

E-mail: vm0011@gmail.com alebo vm001@pobox.sk

Porovnanie jednotlivých krajov SR na základe metódy váženého súčtu poradí

Ľubica Hurbánková

Abstrakt

Na porovnávanie regiónov alebo krajín sa používajú metódy viacrozmerného porovnávania. V príspeku prezentujeme jednu z týchto metód – metódu váženého súčtu poradí na údaje za jednotlivé kraje SR za roky 1997 – 2003. Pred aplikáciou tejto metódy je potrebné stanoviť váhy analyzovaných intenzintých indikátorov. Základným cieľom je usporiadanie regiónov na základe vybraných intnzitných indikátorov – *hrubá pridaná hodnota per capita, účinnosť pracovného vstupu, účinnosť kapitálového vstupu, vybavenosť zamestnancov kapitálom, úplné náklady práce na zamestnanca, podiel počtu zamestnancov a počtu pracujúcich, miera zamestnanosti, miera ekonomickej aktivity* a *podiel pracovného potenciálu na celkovom počte obyvateľov*.

Abstract

Multivariate methods are used for comparison of particular regions or countries. In this paper we present methods of weight definitio of intensity indicators and aplication of method of weighted sum of orders on data of particular regions of Slovak republic for years 1997 – 2003. Basic goal is and ordination of regions based on data of choosed intensity indicators – *gross added value per capita, labour efficiency, capital efficiency, capital – employee ratio, total labour costs per employee, employee efficiency, employees – employed ratio, employment rate, economic activity rate* and *labour potention – population ratio*.

Úvod

Na porovnanie krajov, príp. krajín sa používajú metódy viacrozmerného porovnávania. Hlavným cieľom týchto metód je usporiadanie skúmaných objektov (v našom prípade krajov SR) na základe hodnôt príslušných intenzitných ukazovateľov, a to *hrubá pridaná hodnota na obyvateľa – HPHpc* (podiel hrubej pridanej hodnoty a celkového počtu obyvateľov), *účinnosť pracovného vstupu* – u_L (podiel hrubej pridanej hodnoty a úplných nákladov práce), *účinnosť kapitálového vstupu* – u_K (podiel hrubej pridanej hodnoty a tvorby hrubého fixného kapitálu), *vybavenosť zamestnanca kapitálom – f* (podiel tvorby hrubého fixného kapitálu a počtu zamestnancov), *úplné náklady práce na zamestnanca – m* (podiel úplných nákladov práce a počtu zamestnancov), *produktivita*

zamestnanca – v (podiel hrubej pridanej hodnoty a počtu zamestnancov), *podiel počtu zamestnancov na počte pracujúcich – s_Z, miera zamestnanosti – z* (podiel počtu pracujúcich na počte ekonomicky aktívnych osôb), *miera ekonomickej aktivity – a* (podiel počtu ekonomicky aktívnych osôb na pracovnom potenciáli[1]) *a podiel pracovného potenciálu na celkovom počte obyvateľov - s_P.* Stanovené poradie je ovplyvnené všetkými ukazovateľmi súčasne. Hodnoty použitých intenzitných ukazovateľov sú uvedené v tabuľke 1.

Tab. 1 – Intenzitné ukazovatele za kraje SR (priemerná hodnota za roky 1997 - 2003)

Kraj	*HPHpc*	u_L	u_K	*f*	*m*	*v*	s_Z	*z*	*a*	s_P
Bratislavský	350 684	2,2751	3,2336	192 395	274 818	624 487	0,9261	0,9423	0,7612	0,8412
Trnavský	162 022	2,3225	3,5008	139 395	210 440	489 262	0,9103	0,8295	0,5346	0,8205
Trenčiansky	143 565	2,0002	2,9727	135 756	200 108	403 190	0,9173	0,8847	0,5403	0,8167
Nitriansky	133 535	2,1379	3,2545	130 407	198 572	427 247	0,9108	0,7920	0,5311	0,8239
Žilinský	126 680	1,9553	2,8160	134 933	194 147	381 669	0,9123	0,8383	0,5504	0,7944
Banskobystrický	131 518	1,9022	3,0526	127 169	203 106	391 357	0,9234	0,7880	0,5740	0,8149
Prešovský	95 506	1,7705	2,8280	117 931	187 963	336 194	0,9103	0,7796	0,5243	0,7709
Košický	140 698	1,9376	2,8774	151 946	223 595	438 665	0,9224	0,7740	0,5732	0,7947

Prameň: vlastné výpočty

Pred aplikáciou týchto metód je treba určiť charakter ukazovateľov, t.j. určiť ukazovatele „typu +“ a ukazovatele „typu –“. Všetky intenzitné ukazovatele, s ktorými budeme pracovať, okrem podielu pracovného potenciálu na celkovom počte obyvateľov, sú „typu +“, teda ukazovatele, pri ktorých je žiadúca čo najvyššia hodnota. Podiel pracovného potenciálu na celkovom počte obyvateľov je ukazovateľ „typu – “, je požadovaná čo najnižšia hodnota. Ak by sme požadovali čo najvyššiu hodnotu, znamenali by to úbytok počtu živonarodených detí a tým neskôr aj úbytok pracovného potenciálu. Preto z dlhodobého aspektu požadujeme, aby bol tento podiel čo najnižší.

Stanovenie váh

Ďalší krok, ktorý je potrebné pri aplikácii jednoduchých metód viacrozmerného porovnávnia vykonať, je stanovenie váh daných intenzitných ukazovateľov. Použili sme dva štatistické spôsoby stanovenia váh – a to pomocou variačného koeficientu a štruktúry korelačnej matice.

Pri výpočtoch sa použili:

- Neprevážené intenzitné ukazovatele, kde ukazovateľom nie sú priradené žiadne váhy. Nedostatkom tejto metódy je, že nezohľadňuje dôležitosť ukazovateľa.

[1] Pod pojmom pracovný potenciál rozumieme obyvateľov starších ako 15 rokov.

➢ Váhy vypočítané pomocou variačných koeficientov na základe vzťahu:

$$w_j = \frac{V_j}{\sum_{j=1}^{m} V_j} \tag{1}$$

kde w_j sú váhy j-teho ukazovateľa,

m je počet ukazovateľov,

V_j je variačný koeficient j-tej premennej,

pričom

$$V_j = \frac{s_{x_j}}{\overline{x_j}} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)^2}}{\overline{x_j}} \tag{2}$$

pričom $j=1,2,...,m$

kde s_{xj} je štandardná odchýlka j-teho ukazovateľa,

$\overline{x_j}$ je priemerná hodnota j-teho ukazovateľa,

x_{ij} je hodnota j-teho ukazovateľa v i-tom kraji,

m je počet ukazovateľov.

Tieto váhy sme označili Váhy I. V tabuľke 2 sú uvedené váhy jednotlivých ukazovateľov podľa ich dôležitosti. Najinformatívnejšie sú tie premenné, ktoré majú najväčšiu variabilitu, t.j. najvyšší variačný koeficient.

➢ Váhy získané analýzou štruktúry korelačnej matice na základe vzťahu:

$$w_j = \frac{\left|\sum_{i=1}^{n} r_{ij}\right|}{\sum_{j=1}^{m}\left|\sum_{i=1}^{n} r_{ij}\right|} \tag{3}$$

kde w_j sú váhy j-teho ukazovateľa,

m je počet ukazovateľov,

r_{ij} je párový koeficient korelácie *i*-teho kraja a *j*-teho ukazovateľa,

pričom

$$r_{ij} = \frac{\operatorname{cov} x_i x_j}{s_{x_i} s_{x_j}} \tag{4}$$

kde $cov\ x_i x_j$ je kovariancia medzi *i*-tym krajom a *j*-tym ukazovateľom,

s_{xi} , s_{xj} je štandardná odchýlka *i*-teho kraja a štandardná odchýlka *j*-teho ukazovateľa.

Tieto váhy sme označili Váhy II, sú uvedené v tabuľke 2.

Tab. 2 – Váhy I, váhy II

Ukazovateľ	Variačný koeficient v %	Váhy I – w^1_i	Váhy II – w^2_i
HPH	46,2261	0,3521	0,1198
u_L	8,7663	0,0668	0,0926
u_K	7,4369	0,0566	0,0635
f	15,1314	0,1153	0,1143
m	12,2456	0,0933	0,1163
v	18,8632	0,1437	0,1214
s_Z	0,6710	0,0051	0,0676
z	6,6835	0,0509	0,0938
a	12,7247	0,0969	0,1086
s_P	2,5295	0,0193	0,1021

Prameň: vlastné výpočty podľa hore uvedených vzťahov

Metóda váženého súčtu poradí

Metóda váženého súčtu poradí predstavuje najjednoduchšiu metódu spomedzi metód viacrozmerného porovnávania. Spočíva v tom, že kraje zoradíme podľa každého uvažovaného ukazovateľa. Kraju, ktorého daný ukazovateľ dosahuje najlepšiu hodnotu (najvyššiu pri snahe o maximalizáciu alebo najnižšiu pri snahe o minimalizáciu), priradíme poradie rovnajúce sa počtu krajov v súbore (*n*), kraj s druhou najlepšou hodnotou dostane poradie (*n-1*), až kraj s najhoršou hodnotou daného ukazovateľa poradie (*1*). Tento postup opakujeme pre každú premennú. Ak sa vyskytne rovnaký číselný údaj u viacerých krajov pri jednom ukazovateli, priradíme im rovnaké poradie, určené ako priemer im prislúchajúcich poradí.

Potom integrálny ukazovateľ d_i, podľa ktorého hodnoty usporiadame, získame ako súčet poradí podľa jednotlivých ukazovateľov (s_{ij}) násobených zvolenými váhami ukazovateľov (w_j):

$$d_i = \sum_{j=1}^{m} s_{ij} \cdot w_j \tag{5}$$

pričom $i=1,2,\ldots,n \quad j=1,2,\ldots,m$

Kraj s najvyššou hodnotou integrálneho ukazovateľa (d_i) je vo výslednom poradí na prvom mieste, kraj s druhou najvyššou hodnotou je celkovo druhý, atď.

Výhodou tejto metódy je jej jednoduchosť. Nevýhodou je, že pri jej aplikácii dochádza k čiastočnej strate informácií. Poradie krajov nevystihuje, o koľko je jeden kraj lepší ako druhý. Rovnaké poradie možno získať pri rôzne veľkých kvantitatívnych rozdieloch v ukazovateľoch. Poradie jednotlivých krajov na základe metódy váženého súčtu poradí je uvedené v tabuľke 3.

Tab. 3 – Poradie krajov SR určené metódou váženého súčtu poradí

Intenzitné ukazovatele			**Váhy I**			**Váhy II**		
Por.	**Kraj**	**Integr. ukaz..**	**Por.**	**Kraj**	**Integr. ukaz.**	**Por.**	**Kraj**	**Integr. ukaz.**
1	Bratislavský	70	1	Bratislavský	7,68	1	Bratislavský	7,07
2	Trnavský	54,4	2	Trnavský	6,32	2	Trnavský	5,52
3	Košický	50	3	Košický	5,23	3	Košický	5,17
4	Trenčiansky	48	4	Trenčiansky	5,04	4	Trenčiansky	4,80
5	Banskobystrický	42	5	Nitriansky	4,00	5	Banskobystrický	4,06
6	Nitriansky	39	6	Banskobystrický	3,56	6	Nitriansky	3,78
7	Žilinský	37	7	Žilinský	2,91	7	Žilinský	3,70
8	Prešovský	19,5	8	Prešovský	1,24	8	Prešovský	1,91

Prameň: vlastné výpočty podľa (2.46)

Integrálny ukazovateľ, ktorý sme určili ako jednoduchý aritmetický priemer poradí dal na prvé miesto Bratislavský kraj. Po ňom nasledujú Trnavský, Košický, Trenčiansky, Banskobystrický, Nitriansky, Žilinský a posledný je Prešovský. V podstate sme dostali rovnaké výsledky, či sme použili vážené ukazovatele alebo nevážené. Rozdiel bol iba v tom, ak sme použili váhy I, na piatom mieste bol Nitriansky kraj a na šiestom Banskobystrický, teda si vymenili pozície. Ostatné kraje mali nezmenené poradie.

Záver

Dosiahnuté výsledky pri použití bodovacej metódy by sme mohli zhrnúť nasledovne:

- Na prvých dvoch miestach sa nachádzajú západoslovenské kraje (Bratislavský, Trnavský). Dôvodom je to, že vo všetkých analyzovaných intenzitných ukazovateľoch dosahuje najlepšie hodnoty buď Bratislavský alebo Trnavský kraj. Z toho vyplýva, že najviac rozvinuté v porovnaní s ostatnými kraji SR sú práve tieto kraje. Je to spôsobené najmä tým, že sa tu vytvára podstatná

časť hrubej pridanej hodnoty, pretože svoje sídlo tu má mnoho veľkých podnikov a neustále do tejto oblasti prúdia zahraničné investície na výstavbu nových zahraničných podnikov. Z toho možno usudzovať, že tieto kraje budú i naďalej napredovať.

- Na posledných miestach sa nachádzajú Žilinský a Prešovský kraj. Je to taktiež dôsledkom toho, že tieto kraje dosahujú najnižšie hodnoty analyzovaných intenzitných ukazo-vateľov. Z predchádzajúcich analýz je zrejmé, že v prípade týchto krajov ide o zaostáva-júce kraje. Dôvodov, prečo práve tieto kraje sú v takej situácii, je viacero. Jedným z nich je to, že je tu nedostatok pracovných príležitostí, ľudia odchádzajú z týchto krajov do krajov s lepším uplatnením, čo je dôsledkom toho, že sa tu vyprodukuje oveľa nižšia hrubá pridaná hodnota ako v predchádzajúcej skupine krajov. Tieto kraje priťahujú aj menej zahraničných investorov. Môžeme len dúfať, že táto situácia sa zlepší v Žilinskom kraji po spustení prevádzky zahraničnej firmy KIA v Žiline.

Zoznam použitej literatúry

[1] Hurbánková, Ľ.: Porovnanie regiónov Slovenska pomocou ukazovateľa hrubý domáci produkt na obyvateľa. In: Zborník príspevkov – 13. medzinárodný seminár výpočtová štatistika. Bratislava: SSDS, 2004

[2] Jílek, J.: Metody mezinárodního srovnávaní. Praha: VŠE, 1997

[3] Křovák, J. – Študlar, J.: Metody stanovení vah ukazatelú. In: Statistika 12/1983

[4] Pardelová, R. – Hurbánková, Ľ.: Socio-ekonomické indikátory úrovne hospodárskeho procesu. In: Zborník z medzinárodnej vedeckej konferencie – Vývoj ekonomickej teórie a vzdelávania, uplatnenie a perspektívy v SR. Bratislava: 2005

[5] www.statistics.sk

Autor

Ing. Ľubica Hurbánková

Katedra štatistiky, FHI, EU

Dolnozemská cesta 1/b

852 35 Bratislava

Tel.: 02/672 95 728

e-mail: lu.pod@post.sk

Analýza vplyvu vybraných faktorov na výšku poistných plnení v povinnom zmluvnom poistení SR

Erik ŠOLTÉS – Tatiana ŠOLTÉSOVÁ

Abstract

In the article we detected whether the choice factors (year of production, engine power, region, time) have influence on claim amounts in third party auto insurance. We applied the statistical methods as MANOVA and regression analysis for checking this. Besides we evaluated an influence of this factors on the claim amounts.

Cieľom príspevku je overiť štatistickú významnosť vplyvu vybraných faktorov na výšku poistných škôd v povinnom zmluvnom poistení. Prvoradým zámerom je na základe metód štatistickej indukcie potvrdiť alebo zamietnuť opodstatnenosť klasifikácie poistených vozidiel podľa výkonu motora a územnej príslušnosti.

Vychádzali sme z databázy 45 024 poistiek v priebehu troch rokov v istej nemenovanej poisťovni. Vozidlá sme roztriedili do deviatich tried podľa výkonu motora a ôsmich tried podľa krajov SR. V jednotlivých bunkách kontingenčnej tabuľky (tab. 1) sme zistili počet poistiek (uvedený v zátvorke) a vypočítali priemernú výšku poistných plnení.

Tabuľka 1

Priemerná výška poistných plnení (v Sk) a počet poistených vozidiel

Výkon motora (kW)	*Kraj*								**Spolu**
	BA	BB	KE	NR	PO	TN	TT	ZA	
1. 0 – 11	736 (70)	0 (32)	0 (7)	0 (12)	0 (5)	0 (5)	0 (11)	0 (12)	**335 (154)**
2. 12 – 15	3 584 (11)	0 (4)	0 (1)	30 471 (1)	0 (3)	0 (6)	0 (0)	0 (4)	**2 330 (30)**
3. 16 – 26	0 (23)	77 (16)	1 294 (15)	0 (13)	0 (7)	1 221 (11)	862 (8)	0 (14)	**383 (107)**
4. 27 – 40	2 320 (1 194)	1712 (655)	886 (357)	1 284 (451)	1 564 (253)	1 279 (347)	926 (429)	728 (483)	**1 529 (4 169)**
5. 41 – 55	2 482 (8 234)	2 035 (2 057)	2 348 (1 476)	2 444 (1 798)	1 564 (1 298)	1 807 (1 581)	1 965 (1 607)	1 649 (1 586)	**2 197 (19 637)**
6. 56 – 67	2 505 (2 414)	2 791 (557)	3 034 (418)	1 989 (438)	1 457 (272)	1 360 (440)	1 746 (374)	2 400 (513)	**2 326 (5 426)**
7. 68 – 89	2 691 (4 814)	1 118 (637)	2 601 (530)	1 548 (623)	1 367 (417)	1 875 (760)	1 639 (591)	3 861 (624)	**2 377 (8 996)**
8. 90 – 111	2 642 (1 891)	1 192 (214)	2 626 (241)	1 771 (211)	3 721 (105)	2 149 (288)	2 354 (215)	2 446 (234)	**2 455 (3 399)**
9. 112 -	3675 (1 627)	4 087 (242)	1 114 (140)	1 517 (289)	4 682 (91)	1 939 (267)	5 870 (217)	2 499 (233)	**3 336 (3 106)**
Spolu	**2 627 (20 278)**	**1 998 (4 414)**	**2 272 (3 185)**	**1 995 (3 836)**	**1 718 (2 451)**	**1 747 (3 705)**	**2 018 (3 452)**	**2 096 (3 703)**	**2 274 (45 024)**

Priemerná výška poistných plnení v celom súbore bola 2 274 Sk. Z hľadiska výkonu bola najvyššia priemerná škoda v poslednej výkonnostnej triede nad 112 kW a s výkonom motora klesala, s výnimkou druhej výkonnostnej triedy. V kategórii vozidiel 12 kW až 15 kW

bol rozsah súboru len 30, a preto priemer 2 330 Sk bol ovplyvnený jednou extrémnou hodnotou z Nitrianskeho kraja. Najviac poistených vozidiel bolo s výkonom motora 41 kW až 55 kW. V krajoch dominoval Bratislavský kraj s počtom poistených vozidiel 20 278 a s najvyššou priemernou výškou plnení 2 627 Sk. Na druhej strane, najmenej poistených vozidiel (2 451) a najmenšia priemerná výška poistných plnení (1 718 Sk) boli v Prešovskom kraji.

Okrem týchto dvoch kvalitatívnych faktorov sme uvažovali aj s vplyvom veku vozidla, resp. roku výroby vozidla a s vývojom v čase *t*. Pretože niektoré poistky boli zrušené v priebehu poistného obdobia, na overenie signifikantnosti vplyvu týchto faktorov sme použili váženú viacfaktorovú analýzu rozptylu. Vzhľadom na veľkosť súboru $n = 45\,024$ a výpočtovú náročnosť sme využili aplikáciu programového balíka SAS – Enterprise Guide 3.0. Prvotná analýza (tab. 2) potvrdila štatistickú významnosť modelu ($p < 0,0001$).

Tabuľka 2

Tabuľka analýza rozptylu

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	18	19 322 414 985	1 073 467 499.2	17.71	**<.0001**
Error	45 005	2.7276695E12	60 608 143.722		
Corrected Total	45 023	2.7469919E12			

Na hladine významnosti 0,1 sme overovali nulové hypotézy štatistickej nevýznamnosti vplyvu jednotlivých premenných (*rok výroby*, t, t^2, *Kraj*, *kW*) na poistné škody. Na základe výstupu uvedeného v tab. 3 sme v prípade premenných: t, t^2, *Kraj*, *kW* nulovú hypotézu zamietli ($p < 0,0001$), a teda tieto faktory štatisticky významne ovplyvňujú poistné škody. Významnosť vplyvu veku vozidiel (premenná: *rok výroby*) na poistné plnenia sa na hladine významnosti 0,1 nepotvrdila ($p = 0,125 > \alpha = 0,1$).

Tabuľka 3

Test štatistickej významnosti vplyvu jednotlivých faktorov

Source	DF	Type III SS	Mean Square	F Value	Pr > F
rok výroby	1	142 644 962	142 644 962	2.35	**0.1250**
t	1	3 850 284 800	3 850 284 800	63.53	**<.0001**
t2	1	4 332 396 208	4 332 396 208	71.48	**<.0001**
Kraj	7	3 750 478 278	535 782 611	8.84	**<.0001**
kW	8	5 430 455 866	678 806 983	11.20	**<.0001**

Ďalej nás zaujímalo, medzi ktorými úrovňami faktora *Kraj* ako aj úrovňami faktora *kW* je signifikantný rozdiel v priemerných výškach poistných plnení. Podľa výstupu v tab. 4 sa na hladine významnosti 0,1 potvrdil významný rozdiel ($p < 0,1$) medzi najdrahším Bratislavským krajom a ostatnými krajmi. Na druhej strane, v Prešovskom kraji a Trenčianskom kraji, kde boli zistené najnižšie priemerné hodnoty, sa stredné hodnoty poistných plnení s veľkou pravdepodobnosťou od seba nelíšia ($p = 0,8673 > \alpha = 0,1$).

Tabuľka 4

Matica testov zhody stredných hodnôt

Least Squares Means for effect Kraj Pr > |t| for H0: LSMean(i)=LSMean(j)								
i/j	**BA**	**BB**	**KE**	**NR**	**PO**	**TN**	**TT**	**ZA**
BA		0.0003	0.0673	<.0001	<.0001	<.0001	0.0004	0.0049
BB	0.0003		0.2541	0.6679	0.1360	0.0598	0.8628	0.6253
KE	0.0673	0.2541		0.1332	0.0171	0.0046	0.2159	0.5177
NR	<.0001	0.6679	0.1332		0.2771	0.1582	0.8144	0.3772
PO	<.0001	0.1360	0.0171	0.2771		0.8673	0.2016	0.0627
TN	<.0001	0.0598	0.0046	0.1582	0.8673		0.1079	0.0228
TT	0.0004	0.8628	0.2159	0.8144	0.2016	0.1079		0.5316
ZA	0.0049	0.6253	0.5177	0.3772	0.0627	0.0228	0.5316	

Po zohľadnení výsledkov vyplývajúcich z matice uvedenej v tab. 4, sme kraje medzi ktorými sa s veľkou pravdepodobnosťou potvrdila zhoda stredných hodnôt zlúčili a vytvorili sme novú kategorizáciu:

A – Bratislavský kraj
B – Košický kraj a Žilinský kraj
C – Banskobystrický, Trnavský a Nitriansky kraj
D – Trenčiansky kraj a Prešovský kraj

Tabuľka 5

Matica testov zhody stredných hodnôt

Least Squares Means for effect kW Pr > |t| for H0: LSMean(i)=LSMean(j)									
i/j	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1		0.1458	0.8124	0.0420	0.0038	0.0025	0.0023	0.0020	<.0001
2	0.1458		0.2069	0.4999	0.7609	0.8163	0.8224	0.8502	0.6544
3	0.8124	0.2069		0.1619	0.0351	0.0258	0.0247	0.0220	0.0005
4	0.0420	0.4999	0.1619		0.0001	0.0001	<.0001	0.0002	<.0001
5	0.0038	0.7609	0.0351	0.0001		0.3906	0.2548	0.2583	<.0001
6	0.0025	0.8163	0.0258	0.0001	0.3906		0.9347	0.7193	<.0001
7	0.0023	0.8224	0.0247	<.0001	0.2548	0.9347		0.7475	<.0001
8	0.0020	0.8502	0.0220	0.0002	0.2583	0.7193	0.7475		<.0001
9	<.0001	0.6544	0.0005	<.0001	<.0001	<.0001	<.0001	<.0001	

Podobne sme postupovali v prípade faktora *kW* (tab. 5). Prvé tri výkonnostné triedy vozidiel boli zastúpené len minimálne a navyše s veľkou variabilitou poistných škôd, a tak sa naše očakávania o štatistickej nevýznamnosti rozdielu v stredných hodnotách potvrdili ($p = 0{,}1458 > \alpha = 0{,}1$; $p = 0{,}8124 > \alpha = 0{,}1$; $p = 0{,}2069 > \alpha = 0{,}1$). Na základe porovnania ďalších stredných hodnôt sme sa rozhodli zlúčiť aj iné triedy, čím sme dostali novú klasifikáciu:

1. – 0 kW až 26 kW
2. – 27 kW až 67 kW
3. – 68 kW až 111 kW
4. – 112 kW a viac

Po týchto úpravách sme získali faktor *Kraj** so štyrmi úrovňami: A, B, C, D a faktor *kW** s úrovňami označenými 1., 2., 3. a 4. Priemerné výšky poistných plnení ako aj počty poistených vozidiel (uvedené v zátvorke) sú pre faktory *Kraj** a *kW** zaznamenané v tab. 6.

Tabuľka 6

Priemerná výška poistných plnení (v Sk) a počet poistených vozidiel

Výkon motora (kW)	*Kraj**				**Spolu**
	A	B	C	D	
1. 0 – 26	875 (104)	366 (53)	398 (97)	363 (37)	**558 (291)**
2. 27 – 67	2 471 (11 842)	1 914 (4 833)	2 022 (8 366)	1 604 (4 191)	**2 126 (29 232)**
3. 68 – 111	2 677 (6 705)	3 065 (1 629)	1 518 (2 491)	1 914 (1 570)	**2 398 (12 395)**
4. 112 -	3 675 (1 627)	1 979 (373)	3 611 (748)	2 636 (358)	**3 336 (3 106)**
Spolu	**2 627 (20 278)**	**2 177 (6 888)**	**2 002 (11 702)**	**1 735 (6156)**	**2 274 (45 024)**

Analýza rozptylu (tab. 7) potvrdila štatistickú významnosť modelu ($p < 0{,}0001$) ako aj významnosť vplyvu všetkých uvažovaných premenných (*rok výroby*, t, t^2, *Kraj**, *kW**) na poistné škody (tab. 8).

Tabuľka 7

Tabuľka analýza rozptylu

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	9	18 084 994 521	2 009 443 835.7	33.15	**<.0001**
Error	45 014	2.7289069E12	60 623 515.544		
Corrected Total	45 023	2.7469919E12			

Tabuľka 8

Test štatistickej významnosti vplyvu jednotlivých faktorov

Source	DF	Type III SS	Mean Square	F Value	Pr > F
rok vyroby	1	428 103 387	428 103 387	7.06	**0.0079**
t	1	3 833 588 888	3 833 588 888	63.24	**<.0001**
t2	1	4 327 685 710	4 327 685 710	71.39	**<.0001**
Kraj*	3	3 806 997 654	1 268 999 218	20.93	**<.0001**
kW*	3	4 217 074 951	1 405 691 650	23.19	**<.0001**

Na hladine významnosti 0,1 sa potvrdil signifikantný rozdiel ($p < 0{,}1$) medzi všetkými dvojicami stredných hodnôt poistných plnení na jednotlivých úrovniach faktora *Kraj** (tab. 9). Preukázateľné boli aj rozdiely v stredných hodnotách poistných plnení na jednotlivých úrovniach faktora *kW** (tab. 10).

Tabuľka 9

Matica testov zhody stredných hodnôt

Least Squares Means for effect Kraj* Pr > |t| for H0: LSMean(i)=LSMean(j)				
i/j	**A**	**B**	**C**	**D**
A		0.0017	<.0001	<.0001
B	0.0017		0.0998	0.0010
C	<.0001	0.0998		0.0291
D	<.0001	0.0010	0.0291	

Tabuľka 10

Matica testov zhody stredných hodnôt

Least Squares Means for effect kW* Pr > |t| for H0: LSMean(i)=LSMean(j)				
i/j	**1**	**2**	**3**	**4**
1		0.0029	0.0010	<.0001
2	0.0029		0.0538	<.0001
3	0.0010	0.0538		<.0001
4	<.0001	<.0001	<.0001	

Nakoniec sme vplyv jednotlivých faktorov vyčíslili prostredníctvom odhadov parametrov regresného modelu, ktoré sú uvedené vo výstupe v podobe tab. 11.

Tabuľka 11

Odhad parametrov regresného modelu

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	-45 821.700	B	13 052.127	-3.51	0.0004
rok vyroby	17.294		6.508	2.66	0.0079
t	6 130.458		770.922	7.95	<.0001
t2	-645.982		76.456	-8.45	<.0001
Kraj* A	796.752	B	114.147	6.98	<.0001
Kraj* B	450.426	B	136.747	3.29	0.0010
Kraj* C	268.007	B	122.822	2.18	0.0291
kW* 1	-2 511.683	B	482.360	-5.21	<.0001
kW* 2	-1 129.356	B	147.230	-7.67	<.0001
kW* 3	-964.787	B	156.850	-6.15	<.0001

Keďže všetky regresné koeficienty sú na hladine významnosti štatisticky významné ($p < 0{,}1$), môžeme ich bodové odhady interpretovať:

- Medzi rokom výroby vozidla a výškou poistných plnení je priamy lineárny vzťah, resp. medzi vekom vozidla a výškou poistných plnení je nepriamy lineárny vzťah. S nárastom veku vozidla o rok výška poistných náhrad v priemere klesne o 17,294 Sk.
- Poistné škody v Bratislavskom kraji (*Kraj*A*) sú v porovnaní s Prešovským a Trenčianskym krajom (*Kraj*D*) v priemere o 796,752 Sk vyššie, za predpokladu že porovnávame poistné udalosti vyplatené v rovnakom roku a u vozidiel s rovnakým rokom výroby a z rovnakej výkonnostnej triedy motora. V krajoch zahrnutých v úrovni *Kraj*B* (*Kraj*C*) je tento rozdiel menší o takmer 350 Sk (o takmer 530 Sk), t. j. v Košickom a Žilinskom kraji (v Banskobystrickom, Trnavskom a Nitrianskom kraji) sú poistné plnenia v porovnaní s Prešovským a Trenčianskym krajom (*Kraj*D*) v priemere o 450,426 Sk (o 268,007 Sk) vyššie, ceteris paribus.
- Medzi výkonom motora a výškou poistných plnení je priamy vzťah. Najvyššie poistné plnenia sú u vozidiel s výkonom motora nad 112 kW (*kW*4*), v priemere o 2 511,683 Sk vyššie ako vo výkonnostnej triede *kW*1* (do 26 kW vrátane). Nasleduje trieda *kW*3*, kde je v porovnaní s najvyššou výkonnostnou triedou *kW*4* výška poistných plnení v priemere o 964,787 Sk nižšia. U vozidiel s výkonom motora od 27 kW do 67 kW (*kW*2*) je výška poistných plnení v priemere nižšia o ďalších 167,569 Sk. Tieto tvrdenia platia za predpokladu, že sa jedná o porovnania poistných plnení vyplatených v rovnakom roku, a to u vozidiel s rovnakým rokom výroby a z rovnakého územia (v súlade s klasifikáciou faktora *Kraj**).

Literatúra:

[1] HAIR, J. F – ANDERSOM, R. E. – TATHAM, R. L. – BLACK, W. C.: *Multivariate data analysis*. New York: Maxwell Macmillan International Edition, 1992.
[2] CHAJDIAK, J.: *Štatistické úlohy a ich riešenie v Exceli*. Bratislava: STATIS, 2005.
[3] PACÁKOVÁ, V.: *Aplikovaná poistná štatistika*. Bratislava: ELITA, 2000.

Autori:

Mgr. Erik ŠOLTÉS PhD.
Katedra štatistiky FHI EU
e-mail: soltes@euba.sk
tel.: 02/672 95 717

Mgr. Tatiana ŠOLTÉSOVÁ
Katedra matematiky FHI EU
e-mail: soltesova@euba.sk
tel.: 02/672 95 808

OBSAH

Úvod	1
Hlavné úlohy Slovenskej štatistickej a demografickej spoločnosti na obdobie rokov 2006 - 2010	2
Správa o činnosti Slovenskej štatistickej a demografickej spoločnosti za obdobie rokov 2002 - 2006	4
Stanovy Slovenskej štatistickej a demografickej spoločnosti	6
Chajdiak J., Luha J.: Z histórie Slovenských štatistických konferencií	14
Šipková M.: Hlavné charakteristiky Bratislavského kraja	17
Ondrejka J.: Malacky – dobré miesto pre život	19
Mesto Malacky	20
Luha J.: Podmienky aproximovateľnosti binomického rozdelenia normálnym rozdelením a stanovenie rozsahu výberu	23
Terek M.: Rozhodovanie s viacerými cieľmi a Taguchiho metóda	26
Garaj I.: Jednostranné preberacie plány meraním	32
Garaj I.: Požiadavky na rozsah náhodného výberu jednostranných preberacích plánov meraním	38
Janiga I.: Garaj I.: Necentralne *t*-rozdelenie v odhadovaní minimálnej detegovateľnej hodnoty pomocou lineárnej kalibračnej funkcie s konštantnou smerodajnou odchýlkou	44
Terek M, Hrnčiarová Ľ.: Zisk zo stratifikácie	51
Koróny S.: Zhluková analýza produkčných ukazovateľov stavebných podnikov právnej formy s. r. o.	56
Pastor K.: Váhy polynomických klzavých priemerov	61
Chajdiak J.: Úlohou štatistiky je ukázať svet taký, aký je!	68
Stehlíková B.: Zamestnanosť podľa odvetví ekonomickej činnosti v štátoch EÚ	72
Stankovičová I.: Základné princípy stratégie modelovania	78
Potocký R.: Štatistické analýzy založené na redukcii výberového priestoru	82
Ando M.: Aplikácia klastrovej analýzy pri určovaní bioindikátorov v povodí	85
Vojtková M.: Priestorová klasifikácia priemyselných podnikov podľa efektívnosti	90
Wsólová L., Džupinková Z., Dušinská M.: Genetický polymorfizmus a markery poškodenia DNA	96
Megyesiová S.: Význam analýzy hlavných komponentov pri riešení úloh viacrozmernej štatistiky	101
Glaser – Opitzová H.: Špecifické problémy sezónnej analýzy časových radov	108
Solčánska H.: Metódy na zaistenie konzistencie štvrťročných a ročných časových radov.(Benchmarking)	114
Kaščáková A., Nedelová G.: Analýza kyvadlovej migrácie v okresoch severovýchodného Slovenska	121
Kusendová D.: Vzdialenostné operátory – geoinformačné nástroje priestorovej štatistiky	125
Príhodová I.: Index priemyselnej produkcie	131
Vavrová M.: Datamining	137
Hurbánková Ľ.: Porovnanie jednotlivých krajov SR na základe metódy váženého súčtu poradí	145
Šoltés E., Šoltésová T.: Analýza vplyvu vybraných faktorov na výšku poistných plnení v povinnom zmluvnom poistení SR	151

FORUM STATISTICUM SLOVACUM

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

Redakcia
Miletičova 3
824 67 Bratislava 24
Slovenská republika

Fax
02/63812565

e-mail
chajdiak@statis.biz
Jan.Luha@statistics.sk

Registráciu vykonalo
Ministerstvo kultúry Slovenskej republiky

Registračné číslo
3416/2005

Tematická skupina
B1

Dátum registrácie
22. 7. 2005

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

ISSN 1336-7420

Redakčná rada
RNDr. Peter Mach – *predseda*

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

RNDr. Ján Luha, CSc. – *tajomník*

členovia:
Ing. Mikuláš Cár, CSc.
Ing. Ján Cuper
Ing. Edita Holičková
Doc. RNDr. Ivan Janiga, CSc.
Ing. Anna Janusová
RNDr. PaedDr. Stanislav Katina, PhD.
Prof. RNDr. Jozef Komorník, DrSc.
RNDr. Samuel Koróny
Doc. Ing. Milan Kovačka, CSc.
Doc. RNDr. Bohdan Linda, CSc.
Prof. RNDr. Jozef Mládek, DrSc.
Doc. RNDr. Oľga Nánásiová, CSc.
Doc. RNDr. Karol Pastor, CSc.
Prof. RNDr. Rastislav Potocký, CSc.
Doc. RNDr. Viliam Páleník, PhD.
Ing. Iveta Stankovičová, PhD.
Doc. RNDr. Beata Stehlíková, CSc.
Prof. RNDr. Michal Tkáč, CSc.
Ing. Vladimír Úradníček, PhD.
Ing. Boris Vaňo
Doc. MUDr Anna Volná, CSc., MBA.
Ing. Mária Vojtková, PhD.
Prof. RNDr. Gejza Wimmer, DrSc.
Mgr. Milan Žirko

Ročník
II.

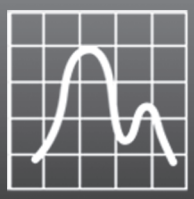
Číslo
1/2006

Cena výtlačku 500 SKK / 20 EUR
Ročné predplatné 1500 SKK / 60 EUR

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



StatSoft®

česká verzia

StatSoft CR, Podbabská 16, CZ-160 00 Praha 6
www.StatSoft.sk, info@StatSoft.cz, tel. +420 233 325 006