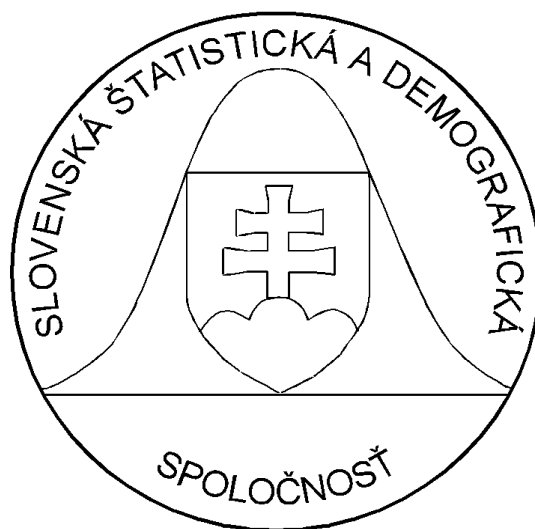


1/2010

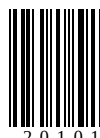
FORUM STATISTICUM SLOVACUM



ISSN 1336-7420



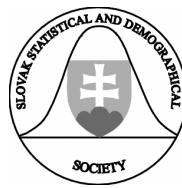
9 771336 742001



20101



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



Naše najbližšie akcie:

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

Pohl'ady na ekonomiku Slovenska 2010

13. 4. 2010, Bratislava, Aula EU

NITRIANSKE ŠTATISTICKÉ DNI 2010

27. - 28. 5. 2010, Nitra

EKOMSTAT 2010, 24. škola štatistiky

tematické zameranie: *Štatistické metódy vo vedecko-výskumnej, odbornej
a hospodárskej praxi.*

30. 5. – 4. 6. 2010, Trenčianske Teplice

FernStat_CZ 2010

23. - 24. 9. 2010, Ústí nad Labem

15. SLOVENSKÁ ŠTATISTICKÁ KONFERENCIA

7. – 8. 10. 2010, Kaskády

19. Medzinárodný seminár VÝPOČTOVÁ ŠTATISTIKA,

2. – 3. 12. 2010, Bratislava, Infostat

Prehliadka prác mladých štatistikov a demografov

2. 12. 2010, Bratislava, Infostat

Regiónálne akcie

priebežne

ÚVOD

Vážené kolegyně, vážení kolegovia,

prvé číslo šiesteho ročníka vedeckého časopisu Slovenskej štatistickej a demografickej spoločnosti (SŠDS) je zostavené z príspevkov, ktoré sú obsahovo orientované v súlade s tematikou 1. Celoslovenskej konferencie medicínskej štatistiky MEDSTAT 2010. Táto akcia sa uskutočnila pod záštitou primátora mesta Ružomberok Ing. Michala Slašťana v dňoch 26. až 28. marca 2010 na Donovaloch.

Akciu z poverenia Výboru SŠDS, zorganizoval:

Programový výbor: Prezident konferencie - Prof. MUDr. Bartko Daniel, MD., PhD., DSc., FAAN, FRSM, FAHA, MUDr. Madarász Štefan, PhD., doc. Ing. Chajdiak Jozef, CSc., RNDr. Luha Ján, CSc., PaedDr. Bunová Barbora, doc. MUDr. Böhmer Daniel, PhD, RNDr. Koróny Samuel, PhD., doc. MUDr. Ozorovský Vojtech, PhD.

a Organizačný výbor:

MUDr. Madarász Štefan, PhD., doc. Ing. Chajdiak Jozef, CSc., PaedDr. Bunová Barbora, RNDr. Luha Ján, CSc., MUDr. Lopeň Peter, MUDr. Midriak Ján, MUDr. Rohal'ová Jana, VEDECKÁ SEKRETÁRKA - PaedDr. BUNOVÁ Barbora.

Na príprave a zostavení tohto čísla participovali: doc. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc. , Ing. Iveta Stankovičová, PhD.

Recenziu príspevkov zabezpečili: doc. Ing. Jozef Chajdiak, CSc., RNDr. Ján Luha, CSc., Ing. Iveta Stankovičová, PhD., RNDr. Samuel Koróny PhD.

Výbor SŠDS

Základné oblasti a úlohy medicínskej štatistiky Basic sections and tasks of medical statistics

Ján Luha

Abstract: In this article we present basic sections and tasks of medical statistics. We define medical statistics or statistics in medicine. We describe sources of medical data and sources of literature concerning with medical statistics.

Key words: medical statistics, epidemiology, biostatistics, genetics, sections of medical statistics, tasks of medical statistics.

Kľúčové slová: medicínska štatistika, epidemiológia, bioštatistika, genetika, oblasti medicínskej štatistiky, úlohy medicínskej štatistiky.

1. Úvod

Medicínska štatistika (alebo štatistika v medicíne) predstavuje veľmi široké spektrum oblastí, úloh, a metód ich riešenia. Prakticky možno v oblasti medicíny využiť všetko, čo poskytuje štatistika a matematická štatistika a naviac využíva špecifické modely vyvinuté napríklad v epidemiológii, bioštatistike, klinických experimentoch, genetike, diagnostike a i.

Túto širokú problematiku možno nájsť v mnohých prácach, časť z nich uvádzame v prehľade literatúry.

Často krát však nájdeme pod pojmom medicínska štatistika, prípadne bioštatistika diela, ktoré okrem názvu a niekoľko málo príkladov z predmetnej oblasti poskytujú určitý súbor štatistických metód. Tieto však možno nájsť v dostupnej štatistickej literatúre lepšie vysvetlené. Pre medicínsky a zdravotnícky výskum sú ale vhodnejšie práce, ktoré poskytujú skutočné riešenia z predmetnej oblasti a tým môžu slúžiť ako návod, či referenčný rámec na korektné štatistické riešenia výskumných a praktických úloh medicínskej štatistiky.

Považujeme za dôležité, aby výskumníci, nielen zberali „dáke“ dáta, ale aby premyslene plánovali medicínske štúdie.

1. Celoslovenská konferencia medicínskej štatistiky MEDSTAT 2010 je tematicky orientovaná na základné úlohy medicínskej štatistiky a veľmi pragmaticky na témy, ktoré sú pre výskumníkov v tejto oblasti podstatné – ako správne zaznamenávať medicínske dáta, aký štatistický softvér môžeme aplikovať na spracovanie medicínskych dát. Významná časť konferencie je venovaná praktickým možnostiam riešenia štatistických úloh pomocou Excelu. Tieto nosné témy sú doplnené prednáškami, ktoré ilustrujú aplikácie štatistiky na konkrétnych úlohách. Budúce konferencie budú venované vybraným témam medicínskej štatistiky aj s prihliadnutím na potreby účastníkov.

2. Definícia medicínskej štatistiky

Termín medicínska štatistika je viac uprednostňovaný vo Veľkej Británii, kým v USA a Kanade skôr termín bioštatistika. Bioštatistika je aj aplikáciou štatistiky na biologické dáta a aj na agrodáta. Špecifická a významná je oblasť epidemiologickej štatistiky, a v nej napríklad analýza prežitia (survival analysis). Významné sú ďalej oblasti genetickej štatistiky,

farmakologickej štatistiky, demografickej štatistiky. Prieniky alebo disjunkcie významných oblastí medicínskej štatistiky, bioštatistiky, epidemiológie, farmakológie a prípadne ďalšie sa rôznia podľa koncepcií a názorov autorov.

Nebudeme sa zaoberať rôznymi koncepciami, prijmeme nasledovnú definíciu medicínskej štatistiky:

Medicínska štatistika je aplikáciou štatistických poznatkov a metód na oblasť medicíny, zdravotníctva a medicínskej praxe.

Medicínski štatistickí majú byť skúsení v teórii aj praxi štatistiky a schopní vybrať vhodné štatistické metódy na analýzu medicínskych dát.

Medicínska štatistika sa spočiatku zaoberala incidenciou, prevalenciou a úmrtiami podľa chorôb, ich komparáciou a rizikovými faktormi, deskripciou zdravotných charakteristík populácie. Dnes je oblasť medicínskej štatistiky značne rozsiahlejšia.

Medicínska štatistika je bázou pre „Evidence based“ medicínu. Umožňuje zhromažďovanie dôkazov a správnu interpretáciu faktov. Vyžaduje od zdravotníckych profesionálov schopnosť správne vyhodnotiť výskumy a získať tým dôkazy pre objektívne závery.

Niektorí autori kritizujú používanie pojmov štatistika s prívlastkami, teda aj medicínska štatistika a uprednostňujú termíny ako štatistika v medicíne, prípadne matematická štatistika v medicíne. Pričom je tiež ťažko vymedziť prieniky, či disjunkcie štatistiky a matematickej štatistiky.

Z pohľadu MEDSTAU budeme pojmy medicínska štatistika a štatistika v medicíne používať duálne v hore definovanom zmysle.

Okrem početných kníh a učebníc sa problematike medicínskej štatistiky venujú aj špecializované vedecké časopisy. Výber z nich je v zozname literatúry.

3. Základné oblasti medicínskej štatistiky

Základné oblasti medicínskej štatistiky prirodzene vyplývajú zo základných úloh medicíny. Preto aj zoznam základných úloh medicínskej štatistiky musí kopírovať základné úlohy medicíny. Zosumarizovať základné oblasti medicíny ale nie je samozrejme úlohou štatistiky alebo štatistika.

Najzákladnejším špecifickým znakom medicíny je to, že predmetom jej skúmania je človek v zdraví i chorobe. Príznačným pre medicínu je poslanie zabrániť chorobným stavom alebo ich s úspechom liečiť. Medicína má vždy do činenia s osudom človeka, ktorý sa domáha pomoci. Lekár preto musí aktívne pristupovať k lekárskeym problémom a konať i vtedy, keď sú teoretické predpoklady ešte kusé. Má povinnosť liečiť aj za veľmi obmedzených znalostí podstaty choroby, kedy je nevyhnutný empirický prístup.

Zoznam medicínskych disciplín je samozrejme značne rozsiahly a nie je predmetom tohto príspevku takýto zoznam predložiť. Abecedne by sme mohli začať akútnou medicínou až po zubné lekárstvo.

Medicínska štatistika využíva všetky metódy štatistiky, pričom mnohé sú špecifickým spôsobom rozvinuté práve v medicínskych aplikáciách. Pomerne ucelená oblasť je Epidemiológia, ktorú niektorí autori považujú za pomerne samostatnú súčasť Medicínskej štatistiky, resp. Bioštatistiky. Najmä v USA prevláda pojem Bioštatistika prakticky ale v oblasti

medicíny, značí to isté ako medicínska štatistika. Ako bolo už spomenuté, prv bioštatistika nachádza uplatnenie aj v iných oblastiach biológie a agrológie, prípadne environmentalistiky a i.

Postupne sa záujmová oblasť medicínskej štatistiky rozvinula do širokých dimenzií. Môže sa stať, že aj profesionál v danej oblasti sa nestretol s niektorou špecifickou oblasťou, ktorá je venovaná medicínskej štatistike. Je to dôsledok veľmi širokého spektra problémov a úloh, ktoré medicínska štatistika rieši.

Základné oblasti medicínskej štatistiky by bolo možné identifikovať z oblastí popísaných v prácach z **Medicínskej štatistiky, Bioštatistiky, Biometriky, Epidemiológie, Genetiky, Diagnostických testov, Klinických výskumov, Klinických experimentov, Zdravotných výskumov, Modelovania chorobnosti, Plánovania zdravotnej starostlivosti, Plánovania biologických experimentov, Laboratórnych vyšetrení a laboratórnych metód, Longitudinálnych štúdií, Analýzy prežitia, Farmakológie, Teratológie, Populačnej štatistiky, Etiológie, Fyziológie, Vital and health statistics a i.**

Ako už bolo spomenuté medicínska štatistika je rozsiahla a tak sa nemusí podariť „vymenovať“ všetky jej základné oblasti – zostane teda stála úloha tento zoznam vytvoriť a dopĺňať.

Treba povedať, že je veľkým rizikom vytvoriť takýto zoznam, autor tohto príspevku uvíta námety na vylepšenie skôr než kritiku.

Zoznam základných oblastí medicínskej štatistiky:

- Výskum distribúcie a determinantov chorôb ľudskej populácie;
- Distribúcia a frekvencia výskytu chorôb;
- Pozorovanie a skúmanie ľudí v ich prirodzenom prostredí;
- Skúmanie rizikových faktorov chorôb;
- Realizácia diagnostických testov;
- Klinické experimenty, ich plánovanie;
- Prirodzená história a vývoj chorôb;
- Štúdium prognostických faktorov, ktoré sú determinanty vývoja chorôb;
- Vyhodnotenie liečenia chorôb;
- Podklady na plánovanie zdravotnej starostlivosti;
- Skúmanie preventívnych faktorov;
- Skúmanie faktorov ovplyvňujúcich danú chorobu;
- Komparácia účinnosti liečebných metód;
- Analýza účinnosti liekov;
- Monitorovanie kvality zdravotnej starostlivosti pacientov;
- Identifikácia ochorení a možných anomálií;
- Určenie najvhodnejšej liečby;
- Skúmanie kvality života pacientov;
- Ochrana verejného zdravia obyvateľov;
- Podpora lekárskeho výskumu
- Porovnanie výsledkov liečebných metód
- Zvyšovanie efektivity;
- Boj s podvodným správaním;
- Predikcie vývoja šírenia určitej chorôb;
- Detekcia faktorov ovplyvňujúcich výskyt určitej nemoci;
- Analýza vplyvu lieku na pacienta;
- Zlepšenie starostlivosti o chorých;

- Analýza prežívania;
- Genetické štúdie, humánna genetika;
- Ekologické štúdie a predpovedanie;
- Skúmanie výživy s ohľadom na zdravie človeka;
- Skúmanie spokojnosti pacientov s liečením;
- Sledovať zdravotný stav populácie (zistiť výskyt ochorení, zistiť distribúciu výskytu ochorení v čase, mieste a priestore, sledovanie (surveillance) incidencie a prevalence určitých chorôb);
- Sledovanie podvodov (zdravotné poisťovne);
- a i.

Na základné oblasti medicínskej štatistiky sa prirodzene viažu základné úlohy medicínskej štatistiky (kapitola 4) a základné štatistické metódy medicínskej štatistiky. V určitom ponímaní by sme mohli oblasti a úlohy medicínskej štatistiky stotožniť.

Ako už bolo niekoľkokrát uvedené je oblasť medicínskej štatistiky rozsiahla, spektrum štatistickým metód prirodzene tiež. A ako už tiež bolo zmienené v medicínskej štatistike sa uplatňujú všetky metódy štatistiky a navyše špecifické metódy a aplikácie vyvinuté priamo v oblasti medicínskej štatistiky. Prehľad štatistických metód medicínskej štatistiky nie je účelom tohto príspevku. Určitý pokus o súpis metód medicínskej štatistiky možno nájsť na stránke: <http://www.animatedsoftware.com/statglos/statglos.htm>.

Další „zoznam“, ktorý je relevantný medicínskej štatistike je zoznam **štatistických znakov**, ktoré sa „vyskytujú“ v medicínskej štatistike a samozrejme, že tieto znaky tvoria taktiež veľmi široké spektrum. Primárne ide o deskriptívne charakteristiky zdravotného stavu populácie. Ide napríklad o charakteristiky narodenia, úmrtia, úroveň (rate) detskej úmrtnosti, incidencia a prevalencia chorôb a ich vývoj v čase a priestore. Štatistické znaky sú skúmané v rozličných medicínskych výskumoch, štúdiách, klinických pokusoch a ich výpočet nie je prakticky dostupný.

4. Základné úlohy medicínskej štatistiky

Základné úlohy medicínskej štatistiky vyplývajú z riešenia úloh v jej základných oblastiach. Tieto úlohy prirodzene súvisia s úlohami, ktoré riešime vo výskume, počnúc od deskripcie javov a procesov, cez rôzne komparácie, po skúmanie závislostí, zákonitosti a predikcie. A samozrejme súvisia aj so štatistickými metódami, ktoré tieto úlohy riešia.

Pri riešení štatistickej úlohy je potrebné a podstatné, na základe poznania ako dáta vznikli, poznania skúmaných štatistických súborov, znakov a identifikácií, správne vybrať adekvátne štatistické metódy. Riešenie tohto problému je možné pomocou štatistických príručiek a vhodného softvéru, pričom pri riešení zložitých štatistických úloh je dobré spolupracovať so štatistikom, ktorý má skúsenosti s riešením úloh daného typu.

Základné úlohy medicínskej štatistiky sú determinované úlohami pri zbere, kontrole a štatistickej analýze dát pochádzajúcich z jej základných oblastí. Tieto úlohy sú principiálne zhodné s úlohami, ktoré štatistika rieši v každej vednej oblasti.

Môžeme teda špecifikovať **všeobecné úlohy, ktoré štatistika obvykle rieši:**

- plán štatistickej úlohy

- definícia záujmového štatistického súboru, štatistických jednotiek a štatistických znakov
- stanovenie rozsahu súboru, resp. výberového súboru (v prípade výberového skúmania)
- zber a záznam dát
- kontrola dát a manipulácia s dátami (napríklad transformácia dát s cieľom vytvoriť odvodené znaky ako napríklad miera výskytu určitej choroby) pred štatistickým spracovaním, riešenie problému chýbajúcich dát, riešenie problému nejednoznačnosti (multiresponse) – prípad znakov s viac možnými hodnotami
- štatistické spracovanie
- vypracovanie reportu a publikovanie.

Pričom samotné štatistické spracovanie rieši úlohy (v závislosti od typov štatistických znakov):

- deskripcia dát a ich prezentácia
- testovanie výskumných hypotéz
- komparácie dvoch, prípadne viac súborov
- skúmanie závislostí/asociácií dvoch a viac znakov
- skúmanie vzťahov medzi štatistickými znakmi
- analýza trendov
- predikcie
- špecifické mnohorozmerné modely
- priestorové modely
- longitudinálne štúdie
- a i.

Najzákladnejšou úlohou medicínskej štatistiky je hľadať korektné štatistické riešenia. Ukazuje sa, že najlepšou cestou ako dosiahnuť takýto cieľ je spolupráca medicínskeho odborníka (odborníkov) so štatistikom (štatistikmi) skúseným (skúsenými) s riešením podobných úloh. Na korektné štatistické riešenia je potrebné:

- pripraviť vhodný design štúdie a či výskumu s dôrazom na medicínsku relevantnosť
- aplikovať vhodné štatistické metódy podľa typov medicínskych dát
- vybrať vhodné štatistické softvéry na štatistické analýzy a transformácie dát
- správne interpretovať výsledky štatistických analýz pre medicínske publikovanie
- komunikovať so skúsenými odborníkmi v medicínskej štatistike pri vypracovávaní publikácií
- sústavne sledovať súčasné a predpokladané trendy v medicínskej štatistike.

5. Zdroje medicínskych dát

Zdrojom medicínskych dát sú medicínske výskumy (observačné štúdie, plánované experimenty, klinické pokusy), charakteristiky ľudskej populácie, lekárske záznamy, výsledky laboratórnych vyšetrení, záznamy pacientov, národné registre chorôb (národné

inštitúcie – centrá na zhromažďovanie medicínskych údajov ako napr. National Center for Health Statistics v USA a u nás Národné centrum zdravotníckych informácií), výsledky lekárskeho vyšetrení, výsledky laboratórnych vyšetrení, klinické experimenty, skúmania kvality života pomocou dotazníkov WHO a i.

Stručne uvedieme základné zdroje medicínskych dát.

Národné registre

Dôležitým zdrojom medicínskych dát sú národné registre. Tieto predstavujú časť veľkého množstva administratívnych zdrojov dát. Značnú časť spravuje Národné centrum zdravotníckych informácií, informácie sú na stránke:

http://www.nczisk.sk/buxus/generate_page.php?page_id=483

Sú to registre:

- **Národný register základných zdravotných údajov**
- **Národný register pacientov s onkologickým ochorením**
- **Národný register pacientov s diabetes mellitus 1. typu**
- **Národný register pacientov s vrodenou chybou srdca**
- **Národný register pacientov so srdcovocievnyim ochorením**
- **Národný register pacientov s cievnyim ochorením mozgu**
- **Národný register pacientov s chronickým ochorením pľúc**
- **Národný register pacientov s vrodenou vývojovou chybou**

Okrem toho sú aj Registre mimo správu NCZI:

- **Národný transplantčný register** www.ncot.sk.
- **Národný register pacientov s tuberkulózou** www.hagy.sk.
- **Národný register pacientov s prenosnými ochoreniami** www.vzbb.sk.
- **Národný artroplastický register** sar.mfn.sk.

Žiaľ, dostupnosť dát z Národných registrov je ale obmedzená a v súčasnosti nie sú k dispozícii bezplatne na výskumné účely.

V príspevku (Luha J., 2010) bližšie charakterizujeme Národný register pacientov s vrodenou vývojovou chybou.

Ďalším zdrojom administratívnych údajov pre medicínsku štatistiku sú :

Údaje medzinárodných a národných štatistických organizácií

World Health Organization (WHO); Eurostat; OSN; National Center for Health Statistics; the Centers for Disease Control and Prevention; EUROCAT; The Agency for Healthcare Research and Quality; The American Public Health Association; The American Cancer Society; Národné štatistické úrady; Štatistický úrad SR; Infostat; ÚIPŠ; Zdravotné poisťovne; Ministerstvo zdravotníctva;

A ďalšie inštitúcie relevantné podľa oblasti skúmania. Vzhľadom na široký záber nepovažujeme za potrebné vytvárať ich zoznam.

Indikátory kvality

Podľa zákona sú poskytovatelia zdravotnej starostlivosti povinný sledovať indikátory kvality. Príklad môžeme nájsť na stránke Všeobecnej zdravotnej poisťovne: <http://www.vszp.sk/showdoc.do?docid=230>.

Indikátory kvality pre rok 2008**1. Poskytovateľ, ktorý poskytuje špecializovanú ambulantnú starostlivosť****Účinnosť a primeranosť zdravotnej starostlivosti**

- A.1.1. Skrining rakoviny krčka maternice
- A.1.2. Manažment chronickej starostlivosti – diabetológia
- A.1.3. Manažment chronickej starostlivosti – pneumológia
- A.1.4. Nepredpokladané prijatia po jednodňovej chirurgii

Výsledky zdravotnej starostlivosti

- A.2.1. Zubný kaz, chýbajúce zuby alebo zuby s výplňou u 12 ročných

2. Poskytovateľ, ktorý poskytuje ústavnú zdravotnú starostlivosť**Efektívnosť využitia zdrojov**

- B.1.1. Denná liečba

Vnímanie starostlivosti pacientom

- B.2.1. Hodnotenie poskytovateľa pacientom
- B.2.2. Počet zrušených elektívnych (plánovaných) operácií
- B.2.3. Riešenie sťažností

Výsledky zdravotnej starostlivosti

- B.3.1. Neodkladné opakované prijatia do ústavnej starostlivosti po prepustení - akútny infarkt myokardu
- B.3.2. Neodkladné opakované prijatia do ústavnej starostlivosti po prepustení - pneumónia
- B.3.3. Úmrtia po zlomenine krčka stehnovej kosti (vek nad 65 rokov)
- B.3.4. Úmrtie na infarkt myokardu po neodkladnom prijatí (35-74 rokov)
- B.3.5. Úmrtie na akútnu mozgovú-cievnu príhodu
- B.3.6. Úmrtie v nemocnici po neodkladnom prijatí na operačný výkon
- B.3.7. Úmrtie v nemocnici po plánovanom prijatí na operačný výkon
- B.3.8. Úmrtie po náhrade bedrového kĺbu

Bezpečnosť pacienta

- B.4.1. Nozokomiálne infekcie
- B.4.2. Infekcie rán

3. Poskytovateľ, ktorý poskytuje všeobecnú ambulantnú zdravotnú starostlivosť**Účinnosť a primeranosť zdravotnej starostlivosti**

- C.1.1. Manažment akútnej starostlivosti
- C.1.2. Manažment chronickej starostlivosti
- C.1.3. Vakcinácia detskej populácie
- C.1.4. Vakcinácia proti chrípke

Výsledky zdravotnej starostlivosti

- C.2.1. Pacienti vyšetrení LSPP
- C.2.2. Pacienti, u ktorých bola vykonaná preventívna prehliadka

Významný zdroj dát pre medicínsku štatistiku je samozrejme Internet.

Zdravotná dokumentácia ukrýva veľa užitočných informácií. Žiaľ, ešte stále je väčšina záznamov vedená iba na papieri, prípadne aj na PC vedená dokumentácia je iba „obdoba“ papierovej dokumentácie, a tak informácie v tejto dokumentácii obsiahnuté nemožno využívať na štatistické spracovanie.

Bolo by žiadúce tieto záznamy spresniť a evidovať podstatné údaje o faktoroch „fungujúcich“ pri diagnózach o ktorých je záznam vedený. Vhodne konštruovaný elektronický záznam poskytuje možnosť ďalšej hlbšej analýzy.

Ďalším zlepšením by mohlo byť sledovanie účinnosti liečby pacientov, prípadne skúmanie kvality života, čo sa už v niektorých ambulanciách deje.

Elektronizácia zdravotných záznamov umožní premeniť cintorín dát na materiál na štatistické spracovanie a zdroj potenciálne cenných informácií.

Zdravotné záznamy lekára, prípadne výsledky laboratórnych vyšetrení, sú dôležitým nástrojom dokumentujúcim stav a priebeh ochorení pacientov. Z papierových protokolov je „dolovanie“ informácií namáhavé a zdĺhavé, pritom aj údaje zo súborov pacientov poskytujú dôležité charakteristiky a ich štatistické spracovanie prináša nové poznatky. Ako ilustráciu uvádzame príklad využitia záznamov pacienta.

Záznamy pacientov

Na konkrétnom príklade, z práce Luha J., Chajdiak J., (2000), ilustrujeme využitie zdravotných záznamov pacientov na štatistické spracovanie

Niektorí pacienti s hypertenziou, či už z popudu ošetrojúceho lekára, alebo z vlastnej iniciatívy si vedú záznamy o krvnom tlaku. Obyčajne si merajú, aj podľa odporúčania predajcov zariadení na meranie krvného tlaku: dátum, čas, systolický tlak (SYS) a diastolický tlak (DIA) a pri meraní Vám ešte zvyknú zmerať aj tep (PULSE). Žiaľ nesledujú sa také faktory, ktoré ovplyvňujú krvný tlak (napríklad: hmotnosť, BMI, nahneval Vás šéf v práci, manžel, či manželka, búrlivá oslava, nedajbože rýchlejší pohyb, vplyv liečby a i.).

V citovanom článku sú vybrané dáta oboch autorov meraných v roku 2000. Stručne uvádzame výber tabuľky zaznamenaných dát a charakterizáciu oboch súborov, tak ako sú uvedené v citovanom článku. Najprv ale typická ukážka záznamu pacienta.

Ukážka záznamu dát pacienta

deň	čas	SYS	DIA	Pulse
30.9.98	19,30	149	98	80
streda	19,34	140	93	---
	19,37	135	88	83
	19,40	140	91	47
	19,43	151	102	81
	19,46	145	95	82
	22,35	137	93	56
	22,37	137	91	83

Ukážka transformácie záznamu vhodnej na štatistické spracovanie

porcis	dátum	deň	čas	sys	dia	pulse	mesiac
1	08-I-2001	pondelok	23,45	130,00	87,00	93,00	1
2	08-I-2001	pondelok	23,47	132,00	86,00	94,00	1
3	08-I-2001	pondelok	23,49	127,00	85,00	89,00	1
4	15-I-2001	pondelok	17,02	145,00	81,00	98,00	1
5	15-I-2001	pondelok	17,04	115,00	83,00	81,00	1

Základné štatistické charakteristiky skúmaných súborov

Za celý súbor J.L.2001

	N	Minimum	Maximum	Mean	Std.Dev.
SYS	632	101,00	161,00	127,71	11,52
DIA	632	66,00	117,00	81,89	6,60
PULSE	632	60,00	100,00	83,88	6,02 5

Za celý súbor J.Ch.2001

	N	Minimum	Maximum	Mean	Std.Dev.
SYS	555	91	173	135,30	14,100
DIA	555	57	120	87,46	12,288
PULSE	555	49	103	66,98	7,753

Korelácie medzi sledovanými premennými

J.L.

Za celý súbor

n=632		SYS	DIA	PULSE
SYS	Pearson	1	,689(**)	-,085(*)
	Sig. (2-t.)	,	,000	,032
DIA	Pearson	,689(**)	1	-,083(*)
	Sig. (2-t.)	,000	,	,037
PULSE	Pearson	-,085(*)	-,083(*)	1
	Sig. (2-t.)	,032	,037	,

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

J.Ch.

Za celý súbor

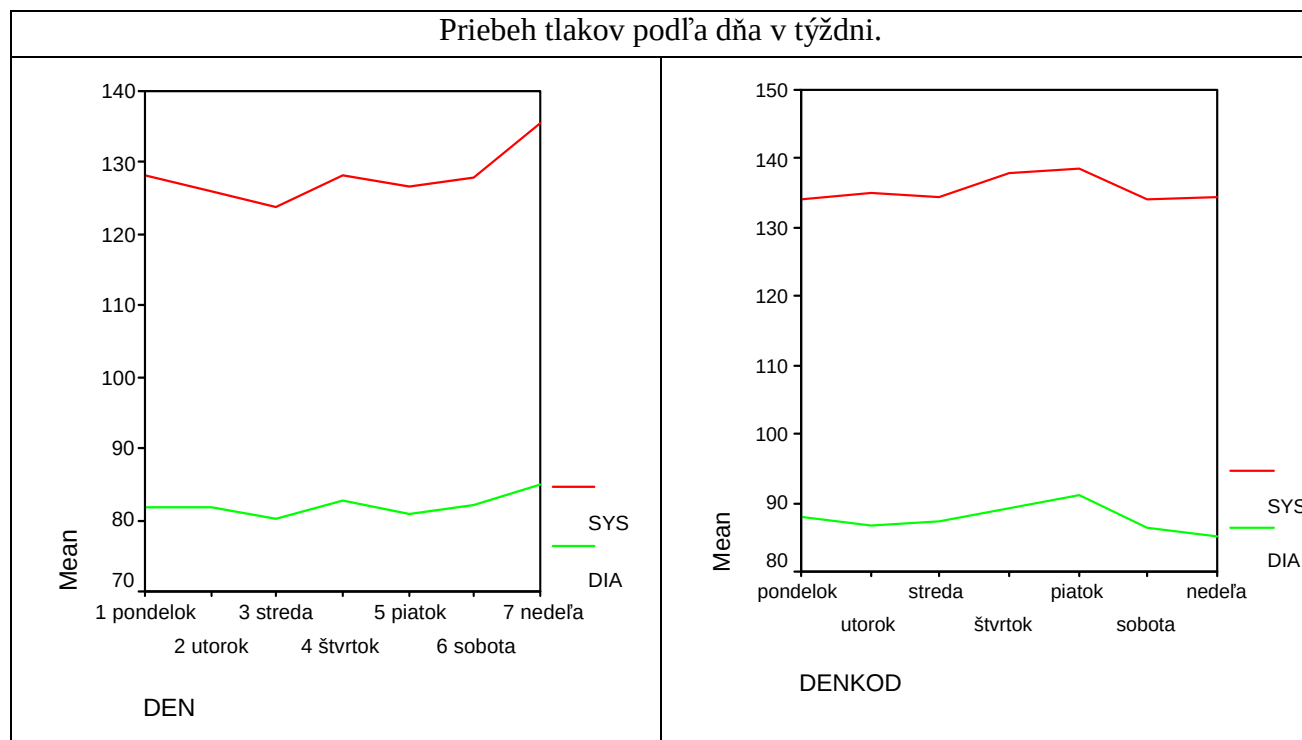
n=555		SYS	DIA	PULSE
SYS	Pearson	1	,819(**)	,085(*)
	Sig. (2-t.)	,	,000	,046
DIA	Pearson	,819(**)	1	,051
	Sig. (2-t.)	,000	,	,232
PULSE	Pearson	,085(*)	,051	1
	Sig. (2-t.)	,046	,232	,

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Z korelačnej analýzy a ako ukazujú aj posledné grafy, existuje pozitívna korelácia medzi systolickým a diastolickým tlakom J.L. 0,689 a J.Ch. 0,819, ale hodnoty korelačného koeficienta medzi tlakom a pulzom sú veľmi malé a nie je ich možné využiť v prípadnom regresnom modelovaní.

Kvôli úspore miesta uvedieme z daného článku ešte zaujímavé grafy priebehu tlakov počas týždňa a komentár autorov.



Z grafickej prezentácie sa pozastavíme pri zaujímavom priebehu tlakov za dni týždňa. Priebeh tlaku jedného z autorov (J.L.) “vykazovalo” zvýšené priemerné hodnoty v nedeľu, čo by mohlo naznačovať zvýšený stres pred nastávajúcim týždňom a môžeme to charakterizovať ako “radosť” z nadchádzajúceho pracovného týždňa. Možno sa domnievať, že podobný model priebehu tlaku má značná časť populácie – v závislosti od “pomerov” v práci. Na rozdiel od uvedeného modelu, priebeh tlaku J.Ch., vykazuje taký charakteristický “stresový” deň až na piatok – čo tiež dačo signalizuje.

Medicínske výskumy a štúdie

Z medicínskych štúdií pochádzajú najvýznamnejšie zdroje medicínskych dát. Nie je prakticky možné vypracovať reprezentatívnu vzorku takýchto štúdií. Na ilustráciu uvedieme tri príklady, viac o je nich v práci (Luha J., 2010).

- Slovenská diabetologická spoločnosť realizovala v roku 2004 špeciálny reprezentatívny celoslovenský výskum „**Prevalencia diabetes mellitus na Slovensku**“ („Skrytá cukrovka na Slovensku“).
- Na istej klinike bol realizovaný **výskum plodnosti žien**. Určité charakteristiky sú vlastné ženám, ale určité zase inému objektu a tým je vajíčko. Pre korektné štatistické spracovanie je dôležité explicitne stanoviť skúmaný súbor, skúmané jednotky a relevantné znaky. Na tomto príklade ilustrujeme dve možnosti, ktoré sú podstatné pre korektné štatistické riešenia.
- Ústav lekárskej biológie, genetiky a klinickej genetiky LF UK a FNŠP Bratislava vyšetruje vzorky potratov, výsledky sú vedené v databáze Abortus. Vzorky zo spontánnych potratov, z indukovaných potratov, zo zamŕknutých potratov a z mŕtvych plodov z bratislavských kliník. Všetky vzorky boli vyšetrované morfológicky a

dlhodobo kultivovali s cieľom dosiahnuť dostatočné množstvo cytogeneticky hodnotiteľných c-metafáz. Databáza v súčasnosti eviduje 3333 vzoriek za obdobie od septembra 1992 do novembra 2009.

6. Literatúra

- [1] Bland M. (2000): Introduction to Medical Statistics, 3rd Edition, Oxford University Press 2000.
- [2] Durham T. A., Turner J. R. (2008): Introduction to Statistics in Pharmaceutical Clinical Trials. PhP 2008.
- [3] Encyklopedia of Biostatistics: http://www.medcalc.be/literature_books.php.
- [4] Feinstein A. R. (2002): Principles of Medical Statistics. Boca Raton London New York Washington, D.C.© 2002 by Chapman & Hall/CRC.
- [5] Fleiss J.L., Levin B ., Myunghee Ch.P.(2003): Statistical Methods for Rates and Proportions. 3ed. Wiley New York 2003.
- [6] Greenberg R.S., Daniels S.R., Flanders W.D., Eley J.W., Boring J.R. III.: Medical Epidemiology. Fourth Edition, McGraw-Hill 2005.
- [7] Handbook of statistics. Vol. 8. Biology and medicine. Editors Rao R.C., Chakraborty R. Elsevier 1991.
- [8] Handbook of statistics. Vol.13. DOE. Editors Ghosh S., R.C., Miller J.P., Rao D.C. Elsevier 1996.
- [9] Handbook of statistics. Vol.18. Bioenvironmental and public health. Editors Sen P. K., Rao R.C. Elsevier 2000.
- [10] Handbook of statistics. Vol. 27 Epidemiology and Medical Statistics. Editors Rao R.C., Miller J.P., Rao D.C. Elsevier 2008.
- [11] Havránek T. (1993): Statistika pro biologické a lékařské vědy. Academia Praha 1993.
- [12] Chajdiak, J.(2003): Štatistika jednoducho. Statis Bratislava 2003, ISBN 808565928-X.
- [13] Chajdiak J. (2005): Štatistické úlohy a ich riešenie v Exceli. STATIS, Bratislava, ISBN 80- 85659-39-5.
- [14] Chajdiak J.(2007): Štatistika v exceli 2007. Statis, Bratislava 2009, ISBN 978-80-85659-49-8.
- [15] Kanji G. K. (2006): 100 Statistical Tests. 3rd Eddition. SAGE 2006.
- [16] Kozlíková K. (2003): Základy spracovanie biomedicínskych meraní I. Asklepios Bratislava 2003. ISBN – 80-7167-064-2.
- [17] Kozlíková K., Martinka J. (2009): Základy spracovanie biomedicínskych meraní II. . Asklepios Bratislava 2009. ISBN – 978-80-7167-137-4.
- [18] Kučera J.: (1989): Populační teratologie. Avicenum Praha 1989.
- [19] Luha J. (1985): Testovanie štatistických hypotéz pri analýze súborov charakterizovaných kvalitatívnymi znakmi. Vydal Odbor Výskumu programov ČST a divákov v SR. Bratislava 1985.
- [20] Luha J., Chajdiak J (2000).: Štatistická analýza dát krvného tlaku? Forum metricum Slovaccum, Tom V. SŠDS Bratislava 2000. ISBN 80-88946-16-6.
- [21] Luha J.(2003a): Skúmanie súboru kvalitatívnych dát. EKOMSTAT 2003. SŠDS Bratislava 2003.
- [22] Luha J.(2003b): Matematickoštatistické aspekty spracovania dotazníkových výskumov. Štatistické metódy vo vedecko-výskumnej práci 2003, SŠDS, Bratislava 2003. ISBN 80- 88946-32-8.
- [23] Luha J. (2006): Štatistické metódy analýzy kvalitatívnych znakov. FORUM STATISTICUM SLOVACUM 2/2006. SŠDS Bratislava 2006. ISSN 1336-7420.

- [24] Luha J.(2007): Kvóťový výber. FORUM STATISTICUM SLOVACUM 1/2007. SŠDS Bratislava 2007. ISSN 1336-7420.
- [25] Luha J. (2008a): Prvotná štatistická analýza kvalitatívnych dát. FORUM STATISTICUM SLOVACUM 2/2008. SŠDS Bratislava 2008. ISSN 1336-7420.
- [26] Luha J. (2008b): Metodologické aspekty zberu a záznamu dát otázok s možnosťou viac odpovedí. FORUM STATISTICUM SLOVACUM 7/2008. SŠDS Bratislava 2008. ISSN 1336-7420.
- [27] Luha J. (2009): Matematicko-štatistické aspekty spracovania dotazníkových výskumov. FORUM STATISTICUM SLOVACUM 3/2009. SŠDS Bratislava 2009. ISSN 1336-7420.
- [28] Luha J. (2010): Metodologické zásady záznamu dát z rozličných oblastí medicíny. FORUM STATISTICUM SLOVACUM 1/2010. SŠDS Bratislava 2010. ISSN 1336-7420.
- [29] Mikulecký M. (1985): Základy biometrie pre experimentálnu a klinickú medicínu. VŠ. skriptá. LF UK Bratislava 1985.
- [30] Riffenburg R. H. (2005): Statistics in Medicine, Second Edition. Academic Press 2005.
- [31] Rosner B. (2000): Fundamentals of Biostatistics. Fifth Edition, Harvard University. Duxbury Press 2000.
- [32] Sheskin D. J. (2004): *Handbook of PARAMETRIC and NONPARAMETRIC STATISTICAL PROCEDURES* .THIRD EDITION, CHAPMAN & HALL/CRC 2004.
- [33] Swinscow T. D. V., Campbell M. J. (1997): Statistics at Square One. Ninth Edition; Copyright BMJ Publishing Group 1997.
- [34] Zvárová J. a kol.: Učebnice projektu EUROMISE: Stochastická genetika; Základy štatistiky; Biomedicínska informatika; Štatistické metódy v epidemiológii; Biomechanika ľudskeho skeletu a umělých náhrad jeho částí.
(<http://ucebnice.euromise.cz/index.php?conn=0§ion=knihy>):
- [35] Woodward M. (2005): Epidemiology. Study design and data analysis. Second Edition. Chapman and Hall/CRC 2005.

7. Časopisecké zdroje

Biometrics

Biometrika

Biostatistics

International Journal of Biostatistics,

The Journal of Agricultural, Biological, and Environmental Statistics

Journal of Biopharmaceutical Statistics

Pharmaceutical Statistics

Statistical Applications in Genetics and Molecular Biology

Statistics in Biopharmaceutical Research

Statistics in Medicine

Turkiye Klinikleri Journal of Biostatistics

8. Internetové zdroje

<http://en.wikipedia.org/wiki/Biostatistics>

http://www.medcalc.be/manual/data_checking.php
<http://www.cdc.gov/epiinfo/>
<http://oli.web.cmu.edu/openlearning/forstudents/freecourses/statistics>
<http://www.medcalc.be/index.php>
<http://statpages.org/>
http://www.hsph.harvard.edu/biostats/links/Biostat_libr.html
<http://www.wiley.com/legacy/wileychi/eob/>
<http://www.wiley.com/legacy/wileychi/eob/articles.html>
<http://www.stahroun.me.cz/interstat/index.htm>
<http://www.statsoft.com/textbook/>
<http://www.spss.com/>
<http://www.texasoft.com/>
<http://www.statsoft.com/textbook/>
<http://davidmlane.com/hyperstat/>
<http://wiki.stat.ucla.edu/socr/index.php/EBook> !!!!!
<http://www.bmj.com/statsbk/>
<http://www.healthnet.sk/texts/statistika/index.htm>
<http://cnx.org/content/col10522/latest/>
<http://www.statisticssolutions.com/>
<http://home.ubalt.edu/ntsbarsh/stat-data/Topics.htm>
<http://www.exploristics.com/>
<http://en.freestatistics.info/data.php>
<http://en.freestatistics.info/stat.php>
<http://www.bmj.com/statsbk/>
<http://www.dkma.dk/1024/visUKLSArtikel.asp?artikelID=1673>
http://www.mdconsult.com/das/book/body/181987038-3/948742465/1657/573.html#4-u1.0-B978-0-323-04743-2..50064-0--cesec1_1605
http://www.mdconsult.com/das/book/body/181987038-4/948743105/1657/573.html#4-u1.0-B978-0-323-04743-2..50064-0_1604
http://www.mdconsult.com/das/book/body/181987038-6/948747765/1393/62.html#4-u1.0-B1-4160-0287-1..50013-6--cesec6_412
http://www.mdconsult.com/das/book/body/181987038-9/948749692/2079/4.html#4-u1.0-B978-1-4160-3143-7..00001-4--s1270_278
<http://www.animatedsoftware.com/statglos/statglos.htm>
<http://www3.interscience.wiley.com/journal/2988/toc?CRETRY=1&SRETRY=0>
<http://www.animatedsoftware.com/statglos/statglos.htm>
<http://www.freestatistics.info/index.php>
<http://statistics.berkeley.edu/~stark/SticiGui/Text/gloss.htm>
<http://www.medicalstatistician.com/>
<http://calcnet.mth.cmich.edu/org/spss/StaProcNonp.htm>
<http://www.intuitivebiostatistics.com/>

Stránky MEDSTAT. Názov MEDSTAT nesú napríklad aj servisné a iné internetové stránky. Pre zaujímavosť uvádzame niektoré z nich.

<http://epp.eurostat.ec.europa.eu/portal/page/portal/medstat/introduction/>
<http://medstatusa.com/default.htm>
<http://www.linkedin.com/companies/thomson-medstat>
<http://www.semide.net/initiatives/medstat>
<http://www.medstatsystems.com/>
[https://leapfrog.medstat.com/\(S\(4n3ufoexjuxs2v45jswzhf45\)\)/index.aspx](https://leapfrog.medstat.com/(S(4n3ufoexjuxs2v45jswzhf45))/index.aspx)

<http://thomsonreuters.com/about/>
<http://medstatcentre.com/>
<http://www.medstatlabsdirect.com/index.html>
<http://www.medstat-finalforum.org/>
<http://www.pragmaticmarketing.com/publications/magazine/4/3/medstat>
http://www.escwa.un.org/information/meetings/editor/Download.asp?table_name=eventDetails&field_name=id&FileID=593
<http://www.encyclopedia.com/doc/1G1-53225356.html>

Adresa autora:

Ján Luha, RNDr., CSc.
Ústav lekárskej biológie, genetiky a klinickej
genetiky LF UK a FNsP
Sasinkova 4, Bratislava
jan.luha@fmed.uniba.sk

**Metodologické zásady záznamu dát z rozličných oblastí medicíny
a zásady ich kontroly**
**Methodological principles of entering data from different sections of
medicine and principles of their checking**

Ján Luha

Abstract: Article dealt with basic principles of entering medical data and principles of their checking. Applications are provided by concrete examples.

Key words: statistical set, statistical units, statistical variables, identifiers, demographical variables, entering data, checking and cleaning of data.

Kľúčové slová: štatistický súbor, štatistické jednotky, štatistické znaky, identifikačné znaky, demografické znaky, záznam dát, dáta, kontrola a čistenie dát.

1. Úvod

Na korektné štatistické spracovanie dát vo všeobecnosti a v medicíne zvlášť, je potrebné dáta správne zaznamenať a taktiež skontrolovať. Žiaľ, až príliš často sa stretávame s problematickými záznamami dát vo forme rozličných tabuliek. Niekedy sú to čiastočné predspracované výsledky, ktoré neumožňujú komplexné štatistické spracovanie danej úlohy. Napriek zdanlivo elementárnemu problému je problém správneho záznamu dát podstatný.

Pri príprave výskumu musíme samozrejme, najprv stanoviť skúmaný súbor a definovať objekty (štatistické jednotky) a skúmané premenné (štatistické znaky). Dôležité je uvedomiť si aký súbor skúmame a čo sú elementárne prvky tohto súboru – štatistické jednotky. Pri štatistických znakoch sa často stretávame s problematikou nejednoznačnosti, lepšie povedané viacznačnosti. Napríklad jeden pacient môže mať viac diagnóz, a aj takéto aspekty je nutné riešiť už pri plánovaní výskumu a pri zázname dát.

Záznam dát možno realizovať rôznymi spôsobmi. Profesionálne agentúry majú systémy umožňujúce elektronický záznam dát napríklad skenovaním, internetovým zisťovaním a pod. Tieto systémy majú už zabudované zásadné princípy platné pre záznam dát a tiež ich kontrolu.

V tomto príspevku sa sústredíme na výklad záznamu dát pomocou Excelu, ktorý má k dispozícii takmer každý, kto má PC.

Zdroje dát pri výskume v medicíne sú rozmanité, niektoré z nich sú uvedené v práci (Luha J., 2010). ***Bolo by vhodné aby sa v projekte elektronizácie zdravotníctva pamätalo aj na elektronizáciu zdravotných záznamov, aby podstatné štatistické znaky bolo možné priamo zaznamenávať do databáz a aby boli zaznamenané aj metadáta o sledovaných štatistických znakoch.***

Motívom výskumníka a aj praktika pre skutočný záujem o štatistické riešenia, vrátane správneho záznamu dát je samozrejme vlastný výskumný problém a tým najmä vlastné dáta výskumníka

2. Štatistický súbor, štatistická jednotka

Pri výskume je dôležité správne stanoviť skúmaný štatistický súbor (záujmový základný súbor (target population) a zodpovedajúci výberový súbor (sample population), prípadne špecifický súbor štúdie (study population) a samozrejme aj jeho elementárne objekty – štatistické jednotky. Dôležité sú samozrejme aj skúmané štatistické znaky, o nich je kapitola 3.

Teoretické definície štatistického súboru a štatistickej jednotky možno nájsť v každej slušnejšej učebnici štatistiky. Na priblíženie niektorých problémov uvedieme príklady.

Príklad 1: Výberový súbor dospelaj populácie SR na výskum výskytu diabetes mellitus

Slovenská diabetologická spoločnosť realizovala v roku 2004 špeciálny reprezentatívny celoslovenský výskum „**Prevalencia diabetes mellitus na Slovensku**“ („Skrytá cukrovka na Slovensku“).

Záujmový základný súbor: Dospelá populácia SR – občania SR vo veku 18 a viac rokov.

Výberový súbor: Reprezentatívny výberový súbor stanoveného rozsahu. Reprezentatívnosť bola sledovaná podľa demografických znakov: pohlavie, vek, vzdelanie, národnosť, veľkostná skupina obce a kraj.

Štatistické jednotky: Občania SR vo veku 18 a viac rokov.

Štatistické znaky: Na výskum bol využitý upravený dotazník WHO, ktorý okrem identifikácie respondenta a demografických znakov obsahoval meritórne znaky podľa definovaných sekcií: S – sekcia o správaní obsahovala otázky na fajčenie a alkohol, D – sekcia s otázkami o konzumácii ovocia a zeleniny, P – sekcia o telesných aktivitách respondentov, H – sekcia o histórii hypertenzie a cukrovky u respondenta, M – sekcia o telesných meraniach ako výška, hmotnosť, obvod pásu, a výsledky vyšetrení o krvnom tlaku, B – sekcia výsledkov biochemických meraní. **Identifikácia záznamov** je jedno-jednoznačné číslo respondenta.

Príklad 2: Skúmanie plodnosti žien s problémom otehotnieť

Na istej klinike bol realizovaný výskum plodnosti žien. Určité charakteristiky sú vlastné ženám, ale určité zase inému objektu a tým je vajíčko. Pre korektné štatistické spracovanie je dôležité explicitne stanoviť skúmaný súbor, skúmané jednotky a relevantné znaky. Na tomto príklade ilustrujeme dve možnosti, ktoré sú podstatné pre korektné štatistické riešenia.

Súbor štúdie 1 (study population): Súbor žien s problémom otehotnieť.

Štatistické jednotky 1: Ženy v štúdiu.

Štatistické znaky 1: Demografické znaky ako je výška, hmotnosť a BMI, výsledky medicínskych vyšetrení vrátane počtu vajíčok (embrií) a ich kvality – tie je ale u určitých súvislostiach nutné skúmať samostatne. **Identifikácia záznamov** je jedno-jednoznačné číslo pacientky.

Súbor štúdie 2: Súbor vajíčok žien s problémom otehotnieť.

Štatistické jednotky 2: Vajíčka (embriá) žien v štúdiu.

Štatistické znaky 2: Tu možno kombinovať znaky zo súboru 1 charakterizujúce jednotky súboru, ale primárne sú štatistické znaky charakterizujúce vlastnosti vajíčok. **Identifikácia záznamov** je jedno-jednoznačné číslo pacientky, prípadne doplnené o identifikáciu vajíčka.

Príklad 3: Národný register vrodenej chyby

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

Záujmový základný súbor: Populácia novorodencov SR s VVCh.

Štatistické znaky: Databáza obsahuje štatistické znaky dané v HVCh, kde okrem identifikácie záznamu, je rok zisťovania, demografické znaky novorodencov aj rodičov (pokiaľ sa podarilo ich zistiť). Výpis z vety novorodenca reprezentuje štatistické znaky, ktoré možno skúmať. Špecifikom tohto súboru, ale aj veľa iných súborov, je viacznačnosť niektorých zisťovaných charakteristík. V tomto prípade sú to charakteristiky výskytu VVCh, keď je zrejmé, že jedinec môže mať aj viac VVCh. NRVVCh dovoľuje záznam do 9 VVCh. Kvôli úplnosti ilustrujeme záznam a teda aj štatistické znaky v NRVVCh.

Názov položky	Popis	
ROK	rok spracovania	
POHLAVIE	pohlavie dieťaťa	1=mužské; 2=ženské; 3=neurčené
TYZDEN_T1	týždeň tehotenstva, v ktorom matka porodila živo, resp. mŕtvonarodené dieťa alebo v ktorom týždni jej bolo vykonané UPT z dôvodu prenatálne zistennej vrodenej chyby plodu	
POC_PLOD	počet plodov v tehotenstve	1=jeden plod; 2=dvojča A; 3=dvojča B; 4=viac ako 2 plody
HMOTNOST	pôrodná hmotnosť	v gramoch
DLZKA	pôrodná dĺžka	v cm
STAV_DIETA	stav dieťaťa	1=mŕtvonarodené; 2=zomrelo do 7.dňa; 3=zomrelo po 7.dni (vrátane); 4=žije
ZACHYT_CH	zachytenosť vrodenej chyby	1=do 7.dňa; 2=po 7.dni (vrátane)
VYS_BIOCHE	uskutočnené prenatálne vyšetrenia - biochemické	1=áno; 0=nie
VYS_CYTOGE	uskutočnené prenatálne vyšetrenia - cytogenetické	1=áno; 0=nie
VYS_ULTRAZ	uskutočnené prenatálne vyšetrenia - ultrazvukové	1=áno; 0=nie
VYS_DNA	uskutočnené prenatálne vyšetrenia - analýza DNA	1=áno; 0=nie
Z_PRENATAL	vrodená chyba plodu zistená prenatálne	1=áno; 0=nie
TYZDEN_T2	týždeň tehotenstva, v ktorom bola prenatálne zistená vrodená chyba	
PRERUS_TEH	prerušenie tehotenstva z dôvodu prenatálne zistennej vrodenej chyby plodu	1=áno; 0=nie
DATUM_P	dátum prerušenia tehotenstva	DDMMRR
RC_MATKY	rodné číslo matky	prvé 6-čísle doplnené 0000
ROD_STAV	rodinný stav	1=slobodná; 2=vydatá; 3=rozvedená; 4=vdova
EA_MATKY	ekonomická aktivita matky - pracovný stav	1=zamestnaná; 2=študujúca; 3=nezamestnaná; 4=príležitostne pracujúca; 5=iná
POC_POROD	predchádzajúce tehotenstvá matky - počet pôrodov	
POC_POTRAT	predchádzajúce tehotenstvá matky - počet spontánnych potratov	
POC_PRERUS	predchádzajúce tehotenstvá matky - počet UPT	
KB1_MATKY	kód okresu trvalého bydliska matky	
BK2_MATKY	druh bydliska bezdomovcov a cudzincov	
KB2_MATKY	kód bydliska bezdomovcov a cudzincov	
R_NAR_OTCA	rok narodenia otca	
EA_OTCA	ekonomická aktivita otca - pracovný stav	1=zamestnaný; 2=študujúci; 3=nezamestnaný; 4=príležitostne pracujúci; 5=iná

MKCH10	diagnózy vrodených chýb dieťaťa podľa MKCH-10 (1.-9.)	číselník MKCH-10 (jedna dg.=4 znaky)
DG_1	diagnózy vrodených chýb dieťaťa podľa MKCH-10 (1.-9.)	číselník MKCH-10 (jedna dg.=4 znaky)
DG_2	diagnózy vrodených chýb dieťaťa podľa MKCH-10 (1.-9.)	číselník MKCH-10 (jedna dg.=4 znaky)
DG_3	diagnózy vrodených chýb dieťaťa podľa MKCH-10 (1.-9.)	číselník MKCH-10 (jedna dg.=4 znaky)
DG_4	diagnózy vrodených chýb dieťaťa podľa MKCH-10 (1.-9.)	číselník MKCH-10 (jedna dg.=4 znaky)
DG_5	diagnózy vrodených chýb dieťaťa podľa MKCH-10 (1.-9.)	číselník MKCH-10 (jedna dg.=4 znaky)
DG_6	diagnózy vrodených chýb dieťaťa podľa MKCH-10 (1.-9.)	číselník MKCH-10 (jedna dg.=4 znaky)
DG_7	diagnózy vrodených chýb dieťaťa podľa MKCH-10 (1.-9.)	číselník MKCH-10 (jedna dg.=4 znaky)
DG_8	diagnózy vrodených chýb dieťaťa podľa MKCH-10 (1.-9.)	číselník MKCH-10 (jedna dg.=4 znaky)
DG_9	diagnózy vrodených chýb dieťaťa podľa MKCH-10 (1.-9.)	číselník MKCH-10 (jedna dg.=4 znaky)
VCH_KOMBIN	vrodená chyba kombinovaná	1=áno; 0=nie
VYSKYT	výskyt vrodenej chyby v príbuzenstve dieťaťa	1=rovnaká vrodená chyba; 2=iná vrodená chyba
DATUM_HL	dátum hlásenia	DDMMRRRR

Príklad 4: Databáza Abortus.

Ústav lekárskej biológie, genetiky a klinickej genetiky LF UK a FNŠP Bratislava vyšetruje vzorky potratov, výsledky sú vedené v databáze Abortus.

Súbor štúdie: Vzorky zo spontánnych potratov, z indukovaných potratov, zo zamŕknutých potratov a z mŕtvych plodov z bratislavských kliník. Všetky vzorky boli vyšetované morfológicky a dlhodobo kultivovali s cieľom dosiahnuť dostatočné množstvo cytogeneticky hodnotiteľných c-metafáz. Databáza v súčasnosti eviduje 3333 vzoriek za obdobie od septembra 1992 do novembra 2009.

Štatistické jednotky: Štatistické jednotky sú vzorky embrií.

Štatistické znaky: Identifikáciu zabezpečuje číslo protokolu. Meritórne znaky sú: Morf_K – kde sa zaznamenáva kód hlavného triedenia morfológie do 11 kategórií; morfológie sú pre významné morfológie bližšie špecifikované podrobnejšími kódmi s textovými popismi; ďalej sú evidované informácie o karyotypoch: Kar_K - zaznamenáva kód hlavného triedenia karyotypov do 11 kategórií a označenie neúspešnej kultivácie, pričom ďalšia špecifikácia je taktiež vedená pre vybrané kategórie karyotypov; vek plodu je uvádzaný v trimestroch a keď bolo možné presnejšie zistiť aj v týždňoch. Ďalej je evidovaný vek matky v rokoch (ak bol zistený) a pohlavie plodu (ak ho bolo možné zistiť).

Príklad 5: Prieskum spokojnosti poistencov VŠZP

Ilustrujeme príklad databázy prostredníctvom metodiky prieskumu podľa <http://www.vszp.sk/showdoc.do?docid=231>.

Jedným z kritérií v rámci **indikátorov kvality** je aj **hodnotenie poskytovateľa zdravotnej starostlivosti samotným pacientom**. Tento indikátor kvality vyjadruje subjektívne hodnotenie poskytovateľa z pohľadu pacientov a je tvorený ako syntetický index hodnotenia poskytovateľa, pokrývajúci minimálne okruhy:

- *hodnotenie celkovej spokojnosti so správaním a starostlivosťou personálu*
- *hodnotenie kvality ubytovania a stravy*
- *hodnotenie spokojnosti so zdravotnou starostlivosťou (subjektívny pocit o úspešnosti liečby)*

Na základe Metodického pokynu Ministerstva zdravotníctva SR č. 20406/2007 z 12. 4. 2007 pre zber indikátorov kvality na hodnotenie poskytovania zdravotnej starostlivosti sa u

poskytovateľa ústavnej zdravotnej starostlivosti získava **hodnotenie poskytovateľa od pacientov pomocou výsledkov reprezentatívneho prieskumu**. Údaje zbiera každá zdravotná poisťovňa samostatne pomocou jednotnej sady otázok vo forme dotazníka. Na zber údajov použije každá zdravotná poisťovňa vlastnú metodiku (anketa, zaslanie dotazníkov poštou,...).

VŠZP v období január 2009 až február 2009 uskutočnila reprezentatívny prieskum náhodne vybraných poistencov, ktorí v priebehu roka 2008 absolvovali liečbu v niektorom zariadení ústavnej zdravotnej starostlivosti na Slovensku. Vybraných poistencov oslovila písomne a požiadala ich, aby vyplnili dotazník, ktorý mapoval úroveň ich spokojnosti s konkrétnym zariadením ústavnej zdravotnej starostlivosti. **Dotazník** obsahoval celkovo štrnásť otázok. Z toho prvých dvanásť otázok bolo zameraných na hodnotenie nemocničného zariadenia z pohľadu pacienta a ich výsledky tvorili súčasť syntetického indexu hodnotenia konkrétneho poskytovateľa zdravotnej starostlivosti. Poistenci sa vyjadrovali zvlášť napr. ku správaniu sa lekárov na oddelení, k starostlivosti personálu, ku kvalite ubytovania a stravy a podobne. Zvyšné dve otázky neboli súčasťou syntetického indexu, i keď mali taktiež hodnotiaci charakter.

Metodika výskumu

Výskum sa realizoval v období od januára 2009 do februára 2009. Na zber dát bola použitá technika selfadministrácie - vyplnenie dotazníka samotným respondentom. Listové zásielky s dotazníkmi boli odoslané viac ako 7000 poistencom VŠZP zo všetkých regiónov SR, ktorí absolvovali ústavnú zdravotnú starostlivosť v roku 2008. Dosiahnutá návratnosť bola na úrovni vyše 32%. Pre potrebu reprezentatívnosti boli sledovanými demografickými znakmi vek a pohlavie.

Prieskum nebol realizovaný u poistencov, hospitalizovaných na psychiatrických, patologických, neonatologických oddeleniach, u poistencov mladších ako 18 rokov a starších ako 70 rokov a u poistencov s inou štátnou príslušnosťou ako slovenská.

Dotazník obsahoval nasledujúcich dvanásť otázok, ktoré boli súčasťou súhrnného indexu:

„Ako ste boli spokojný (á)“:

1. so správaním lekárov na oddelení?
2. s informáciami o vyšetreniach a Vašej chorobe zo strany ošetrojúceho lekára?
3. so správaním sestier na oddelení?
4. so starostlivosťou lekárov na oddelení?
5. so starostlivosťou sestier na oddelení?
6. s kvalitou ubytovania na oddelení?
7. s kvalitou stravy na oddelení?
8. s kvalitou upratovania na oddelení?
9. s poskytnutou zdravotnou starostlivosťou?
10. s informáciami od ošetrojúceho lekára o ďalšom postupe pri domácej liečbe?
11. s informáciami od sestier o ďalšom postupe pri domácej liečbe?
12. so zlepšením Vášho zdravotného stavu po prepustení z nemocnice?

Škála odpovedí bola jednotná pre všetky otázky v oboch dotazníkoch. Poistenci mali možnosť použiť maximálne jednu z nasledujúcich odpovedí:

1. spokojný
2. viac spokojný ako nespokojný
3. ani spokojný ani nespokojný
4. viac nespokojný ako spokojný
5. nespokojný

Na základe odpovedí poistencov Všeobecnej zdravotnej poisťovne hodnotiacich mieru spokojnosti s jednotlivými poskytovateľmi ústavnej zdravotnej starostlivosti, bol vypočítaný súhrnný index za zariadenie ústavnej zdravotnej starostlivosti. Všetky indexy nadobúdali hodnoty z intervalu 0 až 100, pričom platí že 0 reprezentovala najväčšiu nespokojnosť a 100 zase najväčšiu spokojnosť.

3. Štatistické znaky a ich meranie

Na štatistických jednotkách sledujeme určité veličiny, ktoré nazývame štatistické znaky, prípadne premenné alebo náhodné premenné, prípadne položky databázy. Štatistické znaky obvykle delíme na kvalitatívne a kvantitatívne a hlbšie ďalej podľa škál ich merania takto:

- kvalitatívne: dichotomické (nazývané aj binárne); nominálne a ordinálne.
- kvantitatívne: diskkrétne a spojité, pričom spojité ešte môžu mať intervalovú škálu alebo pomerovú škálu (s absolútnou nulou, napr. výška, môže byť vyjadrená v rôznych jednotkách napr. v cm, alebo palcoch, ale nulový bod je absolútny).

Kvantitatívny znaky je pre niektoré analýzy vhodné kategorizovať, napríklad vek skúmaných dospelých osôb kategorizujeme na vekové kategórie: 1= 18-24; 2=25-29; 3= 30-39; 4= 40-49; 5= 50-59; 6= 60 a viac. Ale pri zisťovaní kvantitatívnych znakov je vhodné zaznamenávať ich numericky, kategorizovať ich môžeme pri spracovaní. Ak by sme napríklad vek dospeléj osoby zaznamenali v hore uvedených kategóriách a na analýzu by sme potrebovali iné kategórie, tak by sa mohlo stať, že novú kategorizáciu nemožno urobiť.

Pri kvantitatívnych znakoch, ktoré majú jednotku merania, ako sú napríklad fyzikálne veličiny ako výška, hmotnosť a pod. musí byť jednotka merania jednoznačne definovaná. Nie je prípustné napríklad výšku osôb uvádzať u jednej osoby v metroch a u inej v cm a u ďalšej v mm.

Príklady štatistických znakov sú uvedené v predošlej kapitole.

Z definície štatistického znaku, ako jednoznačnej transformácie:

$$O \Rightarrow H,$$

kde

O označuje množinu objektov (respondentov) a

H označuje množinu „hodnôt“,

dostávame **základné (formálne) štatistické požiadavky, ktoré je potrebné rešpektovať pri konštrukcii škál znakov:**

- jednoznačnosť (zaručuje disjunktný rozklad množiny objektov),
- úplnosť (zaručuje zahrnutie všetkých možných hodnôt).

Úplnosť: škála znaku (otázky) musí byť konštruovaná tak, aby zahŕňala všetky možnosti. Majme napr. znak vzdelanie s kategóriami: základné, stredoškolské bez maturity, stredoškolské s maturitou a vysokoškolské. Ak by sme v dotazníku vynechali čo len jeden variant nemáme úplnú škálu.

Jednoznačnosť (disjunktnosť): Jeden objekt nemôže mať priradenú viac než jednu hodnotu.

Poznámka: Aj medicínskych dátových súboroch sa stretávame so znakmi, ktoré sú viacznačné a preto nie sú štatistické znaky v zmysle definície štatistického znaku. Takéto znaky majú špecifické metódy záznamu a aj spracovania.

4. Zásady prípravy zberu dát

Najčastejšou formou záznamu dát je formulár, napríklad dotazník, anketový lístok; formulár výsledkov vyšetrení a i. Môže mať tlačенú podobu ale aj elektronickú. Okrem spôsobu „papier + pero“ sa využívajú rôzne varianty elektronického zberu dát, napríklad telefonicky, pomocou SMS správ, internetom – buď priamo, alebo pomocou e-mailov, prípadne anketár zaznamenáva odpovede priamo do laptopu a pod.

Nech je už spôsob zberu akýkoľvek, je potrebné dodržiavať určité zásady pri príprave formulára na zber dát a poznať prípadné špecifiká rozličných spôsobov zberu dát.

Uvažujme formulár ako základný nástroj na získanie dát. Dôležitý je vzhľad formulára a pokyny na vyplňanie aby nevznikali zbytočné zdroje chýb od začiatku zberu dát po ich elektronický záznam. Ak ide o populačný výskum pomocou dotazníka, je dôležité aby otázky v dotazníku boli formulované zrozumiteľne, jednoducho, pútavo aby „nútili“ respondenta pravdivo odpovedať. Príbuzné otázky je dobré usporiadať do blokov.

Na overenie zrozumiteľnosti je potrebné otestovať vyplnenie dotazníka pred jeho rozšírením.

Z hľadiska počítačového spracovania súboru je nutné dodržiavať zásady:

Dáta sú namerané (alebo inak získané) hodnoty štatistických znakov na skúmaných objektoch. Je potrebné aby:

- každý objekt mal jedno-jednoznačnú identifikáciu, pritom môžeme okrem identifikácie záznamu (napr. poradové číslo dotazníka) identifikovať aj subjekt, ktorý záznam získal a pod.
- každé pozorovanie musí byť kódované s presnosťou akou bolo realizované meranie a s typom znaku a škálou v akej znak uvažujeme (kvantitatívna (numerická), ordinálna, nominálna, dichotomická). V prípade kvantitatívnych znakov môžeme kódovať s desatinnou bodkou (alebo čiarkou, podľa nastavenia klávesnice). Platný počet cifier musíme rozšíriť o desatinnú bodku.
- pre každú premennú treba rezervovať patričný rozsah, aby bolo možné zaznamenať všetky pozorovania.
- preferujeme numerické kódovanie odpovedí (ich popisy (labele), ako aj názvy premenných (metadáta) môžeme v profesionálnych softvéroch, ako napr. SPSS jednoducho definovať).
- sme mali identifikované, prípadne aj špeciálne okódované, chýbajúce hodnoty.
- bola jednoznačne rozlíšená 0 (nula) pri numerických premenných od chýbajúcich (missing) hodnôt.
- štatistický znak mal jednoznačne pridelené kódy odpovedí, špeciálny prípad predstavujú otázky (znaky) s viacerými možnosťami odpovede (multiresponse).
- kódovanie a záznam multiresponse môžeme realizovať dvomi spôsobmi: Prvý spôsob: okódujeme všetky možnosti odpovedí od 1 po r a na záznam vytvoríme toľko stĺpcov, koľko je maximálny počet dovolených odpovedí, napr. m. Kódy v rozmedzí od 1 po r zaznamenávame do m stĺpcov. Druhý spôsob: na záznam multiresponse otázky pripravíme r, stĺpcov kde r je maximálne možný počet odpovedí a zaznamenáme kód 1 v každom stĺpci pre záznam príslušnej odpovede, ak si túto respondent vybral.

V prípade multiresponse znakov považujeme za chýbajúci údaj, keď sme od daného respondenta nezískali na túto otázku žiadnu odpoveď.

- pri zázname numerickej otázky uprednostňujeme záznam príslušnej numerickej odpovede. Ak je nutné kategorizovať, tak si kategorizáciu vytvoríme pomocou príslušného software.
- pre otvorené otázky je potrebné pripraviť kódovacie kľúče z odpovedí respondentov a zaznamenáme príslušné kódy.
- niekedy môžeme dva alebo niekoľko znakov skombinovať do jedného, bez straty informácie. Uvažujme jednoduchý príklad: „Pohlavie“: 1=muž, 2=žena. Ďalej „Ste bohatý“: 1=áno, 2=nie. Spojená otázka: pohlxbohaty musí mať odpovede, ktoré vzniknú „násobením“ kombinácií oboch otázok: 1=muž,bohatý, 2=muž,nebohatý, 3=žena,bohatá, 4=žena,nebohatá. Skombinovanie môže byť vhodné na redukciu počtu otázok dotazníka.
- okrem identifikácie sú pri populačných prieskumoch dôležitou súčasťou demografické otázky.
- na overenie spoľahlivosti respondentov využívame technické otázky, napríklad pri zvlášť dôležitých otázkach dotazníka.
- pred samotným zisťovaním najme pri rozsiahlych populačných prieskumoch je dôležité otestovať dotazník (formulár na záznam dát) u niekoľkých dobrovoľníkov, prípadne pilotným zisťovaním.

Všeobecné zásady premietneme do vzoru formulára na záznam dát a budeme ich ilustrovať na konkrétnych príkladoch.

5. Formulár na záznam dát

Je veľmi dôležitým nástrojom zberu dát, pričom jeho formu prispôbujeme spôsobu ich zberu v terénnej fáze prieskumu. Podstatné atribúty spočívajú v exaktnej príprave formulára, rozdiely podľa spôsobu zberu dát v tomto príspevku neskúmame.

Prípravu záznamu dát budeme ilustrovať na príklade dotazníka, ktorý bol použitý na jednej univerzite na ilustráciu záznamu dát a ich kontroly a čistenia.

Dotazník obsahuje všetky základné typy škál štatistických znakov: nominálnu, ordinálnu a numerickej a tiež základné typy otázok: uzatvorené, otvorená a tiež s viac možnými odpoveďami.

LUST		AGENCIÁ VÝSKUMU TRHU A VEREJNEJ MIENKY	
Kód respondenta (porcis):			
		DOTAZNÍK “MEDSTAT 2010“	Číslo protokolu: Mxx/2010
1	01	-	áno raz
	02	-	áno viackrát
	03	-	nie

Na otázku č. 2 odpovedajú respondenti, ktorí na otázku č. 1 uviedli kódy 01 (áno raz) alebo 02 (áno viackrát).

2		Ak áno, čo bolo obsahom dotazníka?
3		Čo si predstavujete pod pojmom štatistická analýza dát ?
4	01 - 02 - 03 - 04 - 05 -	Ste za zavedenie EURA v SR? - rozhodne áno - viac áno ako nie - viac nie ako áno - rozhodne nie - iná odpoveď (uved'te)
5	01 - 02 - 03 - 04 - 05 - 06 -	Povedzte, prosím, či ste čítali učebnicu „Žvábek J.: Výuka statistiky na internete. GRADA, Praha, 2006“? - áno, už 1. vydanie - áno, 2. vydanie - iba som ju prelistoval - nie, ale už ju zháňam - nepoznám tú knihu - iná odpoveď (uved'te).....
6	01 - 02 - 03 - 04 - 05 - 06 - 07 - 08 - 09 - 10 - 11 -	Priznajte sa, ktorú stranu by ste volili? - Komunistickú stranu Slovenska (KSS) - Kresťanskodemokratické hnutie (KDH) - Ľudovú stranu – Hnutie za demokratické Slovensko (ĽS-HZDS) - Slovenskú demokratickú a kresťanskú úniu -Demokratickú stranu (SDKÚ-DS) - Slovenskú národnú stranu (SNS) - Smer – sociálnu demokraciu (SMER) - Stranu maďarskej koalície (SMK) - Stranu Most _Híd - inú politickú stranu, hnutie, koalíciu (uved'te)..... - nezúčastnil by som sa volieb - neviem
7		Ktorému súčasnému politikovi aspoň trochu dôverujete? (Možno uviesť tri odpovede)
8	01 - 02 -	Pohlavie: - muž - žena

9		Uved'te prosím, Váš vek:
----------	--	--

10	01 - 02 - 03 -	Národnosť: slovenská maďarská iná (uved'te)
-----------	----------------------	---

11	01 - 02 - 03 - 04 - 05 -	Ako hodnotíte úroveň stravy vo Vašej jedálni? veľmi dobrá dobrá ani dobrá ani zlá zlá veľmi zlá
-----------	--------------------------------------	---

12	01 - 02 - 03 - 04 - 05 -	Veríte výskumom verejnej mienky? rozhodne áno viac áno ako nie viac nie ako áno rozhodne nie iná odpoveď (uved'te)
-----------	--------------------------------------	--

13	01 - 02 - 03 - 04 -	Páči sa Vám študovať na Vašej univerzite? áno nie iná odpoveď (uved'te)..... neviem
-----------	------------------------------	--

14	01 - 02 - 03 - 04 - 05 -	Pokladáte tento dotazník za zaujímavý? rozhodne áno viac áno ako nie viac nie ako áno rozhodne nie iná odpoveď (uved'te)
-----------	--------------------------------------	--

Na záznam dát pripravíme elektronický formulár v Exceli. Pre náš príklad má vzhľad:

porcis	o1	o2a	o2b	o2c	o3a	o3b	o4	o5	o6	o7a	o7b	o7c	o8	o9	o10	o11	o12	o13	o14
--------	----	-----	-----	-----	-----	-----	----	----	----	-----	-----	-----	----	----	-----	-----	-----	-----	-----

Ako správne zaznamenať dáta budeme ilustrovať v kapitole 7, predtým vysvetlíme špecifiká záznamu multiresponse znakov.

6. Špecifické aspekty záznamu znakov s viacerými možnosťami odpovede (multiresponse)

Multiresponse otázky môžu mať vopred stanovenú škálu odpovedí alebo môžu byť otvorené, teda respondent formuluje odpovede sám. V druhom prípade je potrebné pred záznamom dať vytvoriť kódovací kľúč, kde zhľukujeme určité „podobné“ typy odpovedí a priradíme im kód a názov, ktorý tieto odpovede reprezentuje. Dôležité je pred záznamom dať mať k dispozícii „finálnu“ škálu odpovedí (kódovací kľúč).

Otázka s možnosťou viacerých odpovedí (multiresponse question)

Uvažujme otázku Z s množinou možných odpovedí $\{1, 2, \dots, r\}$ a možnosťou dať maximálne m odpovedí, kde: $m \leq r$.

Z	1	2	...	r	Σ
abs. počet. odpovedí	K_1	K_2	...	K_r	K
rel. počet. vzhľadom na počet respondentov	P_1	P_2	...	P_r	P
rel. počet. vzhľadom na počet odpovedí	p_1	p_2	...	p_r	1

Kde K_i je počet odpovedí s hodnotou i , n je počet odpovedajúcich respondentov, K je počet všetkých odpovedí a kde $P_i = K_i/n$, $p_i = K_i/K$.

Platí: $\sum_i K_i = K$, $\sum_i P_i = P$ a $\sum_i p_i = 1$, pričom $P \geq 1$ a $K \leq m \cdot n$.

Príklad: Otázka: Ktorý z politikov pôsobiach na Slovensku má v súčasnosti vašu dôveru? Možno uviesť najviac tri odpovede. (Príklad je zo spojeného súboru štyroch výskumov ÚVVM, ktorý je využitý aj v práci [7].)

Case Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
\$polit(a)	4347	97,3%	122	2,7%	4469	100,0%

a Group

\$polit Frequencies

		Responses		Percent of Cases
		N	Percent	
\$polit(a)	1 žiadnemu	1094	16,2%	25,2%
	2 M.Dzurinda	212	3,1%	4,9%
	3 I.Gašparovič	727	10,8%	16,7%
	4 R.Fico	1467	21,7%	33,7%
	5 V.Mečiar	341	5,0%	7,8%
	6 M.Flašíková-Beňová	58	,9%	1,3%
	7 J.Slota	469	6,9%	10,8%
	8 B.Bugár	235	3,5%	5,4%
	9 V.Tomanová	90	1,3%	2,1%
	10 D.Čaplovič	69	1,0%	1,6%
	11 J.Počiatek	97	1,4%	2,2%
	12 V.Palko	39	,6%	,9%
	13 F.Mikloško	56	,8%	1,3%
	14 P.Hrušovský	139	2,1%	3,2%
	15 D.Lipšic	94	1,4%	2,2%
	16 E.Kukan	27	,4%	,6%

17	P.Paška	60	,9%	1,4%
18	I.Mikloš	175	2,6%	4,0%
19	A.Belousovová	72	1,1%	1,7%
20	I.Radičová	349	5,2%	8,0%
21	V.Veteška	18	,3%	,4%
22	R.Kaliňák	420	6,2%	9,7%
23	Z.Martináková	24	,4%	,6%
24	P.Csáky	86	1,3%	2,0%
25	iní	335	5,0%	7,7%
Total		6753	100,0%	155,3%

a Group

Záznam otázok s viacerými možnosťami odpovede

Kódovanie a záznam multiresponse môžeme realizovať dvomi spôsobmi:

Prvý spôsob: okódujeme všetky možnosti odpovedí od 1 po r a na záznam vytvoríme toľko stĺpcov, koľko je maximálny počet dovolených odpovedí, napr. m. Kódy v rozmedzí od 1 po r zaznamenávame do m stĺpcov.

Príklad: Maximálny počet dovolených odpovedí:

Uvažujme ilustratívnu otázku kde počet odpovedí $r=25$ a maximálny počet možností odpovedať $m=3$. Prvý spôsob záznamu redukuje počet stĺpcov. Záznam, možno realizovať pomocou tabuľky:

porcis	otazka_a	otazka_b	otazka_c
1			
2			
3			
.			
.			

Chýbajúci záznam pre multiresponse otázku identifikujeme, ak v príslušnom riadku nie sú odpovede vo všetkých stĺpcoch.

Tento spôsob záznamu **umožňuje** zaznamenať v jednom riadku jeden kód viackrát, čo pri rekódovaní pomocou transformácií na systém jedna/nula spôsobuje vypadnutie „redundantných“ hodnôt a tým aj odlišné výsledky v štatistických analýzach. V ilustratívnom príklade je kód odpovede 25=iní politici zaznamenaný potenciálne viackrát. Naozaj sa vyskytol v kombinácii prvého a druhého stĺpca 1 krát a 4 krát v kombinácii druhého a tretieho stĺpca. Tento problém sa priamo vo výsledkoch, ktoré produkuje SPSS neprejavuje, iba ak potrebujeme túto premennú rekódovať, prípadne transformovať na 25 indikátorových premenných, ktoré skúmajú elementárne javy vyjadrené v jednotlivých odpovediach v škále otázky.

Druhý spôsob: Na záznam multiresponse otázky pripravíme r, stĺpcov kde r je maximálne možný počet odpovedí a zaznamenáme kód 1 v každom stĺpci pre záznam príslušnej odpovede, ak si túto respondent vybral, a kód 0, keď si túto možnosť nevybral (takto sú definované indikátorové premenné). Aj v tomto prípade **chýbajúci údaj** znamená, že sme od daného respondenta nezískali na túto otázku žiadnu odpoveď. Môžeme použiť aj **modifikáciu**: v i-tom stĺpci pre i-tu odpoveď zaznamenáme kód i, ak respondent volil túto odpoveď a kód 0, keď nie. Pre štatistické vyhodnotenie je to ale menej výhodné ako pomocou indikátorových premenných.

Príklad: Maximálny počet odpovedí:

Druhý spôsob záznamu multiresponse otázok môže mať určitú výhodu, pretože zaznamenáva všetky skúmané elementárne javy, ale má menšiu nevýhodu vo viac stĺpcoch

potrebných na záznam dát. Tento spôsob pôsobí určité problémy v prípade otvorených multiresponse otázok, ako napríklad o dôveryhodnosti politikov, pretože nepoznáme vopred maximálny počet odpovedí a kvôli potenciálnej potrebe použiť kód pre „zvyškové“ odpovede typu „iní“ a pod. Odpovede 1 alebo 0 „zapíšeme“ do príslušného stĺpca tabuľky:

porcis	p_01	p_02	p_03	p_04	p_05	p_06	p_07	p_08	p_08	...	...	...	p_23	p_24	p_25
1															
2															
3															
.															
.															

Kvôli úspore miesta neuvádzame konkrétny výstup pre druhý spôsob záznamu multiresponse otázky. Výsledok je samozrejme rovnaký.

Poznámka: Ak z nejakých dôvodov, napríklad kvôli lepšej prehľadnosti výstupov, potrebujeme zmeniť kódovací kľúč otvorenej multiresponse otázky, musíme skúmať či neporušíme konzistenciu výstupov pri rôznych kódovacích kľúčoch. To môže nastať ak iným spôsobom preskupíme odpovede a teda inak definujeme „podobné“ zhľuky odpovedí. V takomto prípade nemôžeme vytvoriť nové premenné napríklad rekódovaním už zaznamenaných údajov, ale musíme ich znova zaznamenať, čo je pri väčšom rozsahu výberového súboru značne prácne.

Indikátorové premenné otázok s viacerými možnosťami odpovede

Pri transformáciách multiresponse otázok na indikátorové premenné (*nadobúdajú hodnotu 1, ak respondent uviedol v odpovedi daného politika, inak 0*), napríklad pri otázke o dôveryhodnosti politikov, musíme najprv identifikovať chýbajúce odpovede. Kódovanie chýbajúcich hodnôt musí byť realizované takým kódom, ktorý sa nemôže vyskytnúť ako „hodnota“ odpovede na danú otázku. Uvažujme prvý spôsob záznamu multiresponse otázok prezentovaný v 3. kapitole. Musíme rešpektovať kódovanie chýbajúcich hodnôt v tomto zázname, aby sme správne definovali chýbajúce hodnoty indikátorových premenných.

V ilustratívnom príklade sú v pôvodnom zázname *chýbajúce hodnoty označené kódom 0*.

Príklad:

*Najprv zabezpečíme aby hodnoty 0 deklarované ako chýbajúce hodnoty sa pre účely transformácie zmenili na „reálne“ nuly.

Potom vytvoríme pre transformované premenné patričné chýbajúce hodnoty. Keďže pracujeme s dvojcifernými kódmi, označili sme chýbajúce hodnoty indikátorových premenných kódom 99.

Na záver definujeme nuly (označujú „neprítomnosť“ politika v odpovedi respondenta) a znovu označíme chýbajúce hodnoty ako systémové „missingy“.

***korektné definovanie chýbajúcich hodnôt pri multiresponse. Kvôli úspore miesta ilustrujeme syntax iba na niekoľkých politikoch.**

```
MISSING VALUES polit_a TO polit_c ().
```

```
IF (polit_a = 0 & polit_b = 0 & polit_c = 0) ziadny = 99 .
```

```
IF (polit_a = 0 & polit_b = 0 & polit_c = 0) dzurinda = 99 .
```

```
IF (polit_a = 0 & polit_b = 0 & polit_c = 0) gasparovic = 99 .
```

```
IF (polit_a = 0 & polit_b = 0 & polit_c = 0) fico = 99 .
```

```
IF (polit_a = 0 & polit_b = 0 & polit_c = 0) ini = 99 .
```

***definovanie prítomnosti daného politika v odpovedi respondentov kódom 1.**

```
IF (polit_a = 1 or polit_b = 1 or polit_c = 1) ziadny = 1 .
```

```
IF (polit_a = 2 or polit_b = 2 or polit_c = 2) dzurinda = 1 .
```

IF (polit_a = 3 or polit_b = 3 or polit_c = 3) gasparovic = 1 .

IF (polit_a = 4 or polit_b = 4 or polit_c = 4) fico = 1 .

IF (polit_a = 4 or polit_b = 4 or polit_c = 4) ini = 1 .

EXECUTE.

***definovanie neprítomnosti politika v odpovediach respondentov kódom 0 a označenie chýbajúcich hodnôt.**

RECODE

 ziadny to ini (SYSMIS=0) (99=SYSMIS) .

Výsledkom sú indikátorové premenné reprezentujúce „prítomnosť“ daného politika. Ako sme už uviedli prv, pre iných politikov je výnimkou kód 25 a teda premenná **ini**, kde sa po transformácii už nemôžu vyskytnúť prípadné multiplicity tohto kódu. Ilustrujeme to vo výstupe indikátorovej premennej, kde je percento jej výskytu 7,4%, hoci reálne bolo 7,7%.

ini		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	4025	86,7	92,6	92,6
	1	322	6,9	7,4	100,0
	Total	4347	93,6	100,0	
Missing	System	295	6,4		
Total		4642	100,0		

Tento „nedostatok“ sa dá odstrániť rozšírením kódovacieho kľúča. Keďže takáto zmena neovplyvňuje výsledky „skúmaných“ politikov a rozšírenie o ďalších politikov s malým „výskytom“ nie je až také zaujímavé, tak sa zvyčajne od rozširovania rozsahu kódov upúšťa.

Úprava dátovej matice na spracovanie multiresponse otázky

Ďalší spôsob riešenia spracovania multiresponse otázok. ***Táto úprava umožňuje analýzu „multiresponse“ otázky z bázy podľa počtu odpovedí a nie podľa počtu respondentov, preto pri prípadných ďalších analýzach je nutné túto úpravu vziať do úvahy, pretože nie všetky štatistické analýzy spojené aj s inými premennými takto upraveného súboru možno korektne realizovať.***

Úpravu ilustrujeme na príklade otázky o dôveryhodnosti politikov. V tomto prípade máme maximálny počet odpovedí $r=3$. Dátovú maticu skopírujem trikrát „pod seba“ a tri stĺpce záznamu premennej polit_a, polit_b a polit_c „nastavíme postupne „pod seba“. Schématické vyjadrenie vysvetlí celú manipuláciu s dátami:

Pôvodný dátový súbor:

pagina	o1	polit_a	polit_b	polit_c	o2	...	om
1							
2							
...							
n							

skopírujeme trikrát ($r=3$) „pod seba“, pričom pri premennej polit kopírujeme postupne „pod seba“ hodnoty stĺpcov polit_a, poli_b a polit_c:

pagina	o1	polit	o2	...	om
1					
2					
...					

n					
1					
2					
...					
n					
1					
2					
...					
n					

7. Ako správne zaznamenať dáta

Budeme sa zaoberať teda záznamom dát, ktoré máme získané na papierových formulároch (napríklad dotazníky na populačné výskumy, špecifické formuláre pre záznam dát podľa zdroja dát), ktoré potom zaznamenáme do elektronickej podoby.

Skenovanie, internetové, e-mailové a ďalšie spôsoby priameho elektronického zberu dát sú výhodné, keď máme takúto techniku a softvér k dispozícii. Vtedy máme „k dispozícii“ aj odborníkov na realizáciu takéhoto zberu dát. Zásady správneho záznamu však platia aj pre priamy elektronický zber dát.

V kapitole 5 sme už pripravili formulár na elektronický záznam dát.

Pri zložitejších záznamových formulároch potrebujeme explicitne definovať dáta a aj ich popisy, čiže metadáta. Keďže metadáta, až na otvorené a polootvorené otázky, sú v dotazníku nie je nutné ich inak definovať. V prípade, keď nemáme na záznam dát pripravený dotazník, je potrebné exaktne definovať kódy otázok a odpovedí a vytvoriť **príručku kódov** (metadát).

Vo všeobecných zásadách pre záznam dát sa zaoberáme špeciálne chýbajúcimi (**missing**) údajmi. Je dôležité aby chýbajúce dáta boli správne identifikované a „nezmiešali“ sa i inými kódmi. Často sa stáva, že sa „zmiešajú hodnoty 0 a chýbajúce dáta a potom, býva veľmi ťažké realizovať korektné štatistické analýzy. V našom príklade identifikujeme chýbajúce dáta jednoducho tak, že nie je žiaden záznam. V prípade potreby sú chýbajúcim hodnotám pridelené kódy tak, aby sa nemohli „zmiešať“ s možnými hodnotami daného znaku.

V ilustračnom príklade sme v excelovskom záznamovom formulári vytvorili tri listy:

- formulár – prázdny formulár na záznam dát, jeho vzhľad je uvedený hore,
- metadáta – na ilustráciu ako vytvoriť príručku kódov, v danom príklade je však potrebné ešte metadáta – príručku kódov – doplniť o kódovacie kľúče otvorených a polootvorených otázok,
- dáta – kde sme už vyplnili pripravený formulár, na ilustráciu uvádzame dátovú tabuľku, ktorú vyplnili študenti jednej univerzity, odpovedalo 37 študentov.

Po skontrolovaní vyplnených dotazníkov nasleduje ich záznam do elektronickej podoby pre štatistické vyhodnotenie. Pri internetovom výskume je záznam realizovaný do elektronickej databázy a odpadá „ručné“ zaznamenávanie dát. Výskumné agentúry používajú na záznam dát napríklad skenovanie.

Na záznam dát je výhodné používať číselné kódy odpovedí. V našom dotazníku sú tri otvorené otázky, pre ktoré je potrebné pred záznamom dát vypracovať kódovacie kľúče – čo znamená – zoskupiť príbuzné odpovede a ich názvom priradiť číselné kódy. Ostatné otázky dotazníka majú pripravené číselné kódy odpovedí. Na číslo dotazníka sme vyhradili štyri políčka, meno respondenta zaznamenávame iba výnimočne, keď je to pre výskum dôležité.

V rámci prípravy záznamu dát sme zapísali odpovede na otvorené, poloopené otázky ako aj poznámky respondentov. Uvádzame ich v prehľadnej tabuľke.

Otvorené otázky, popootvorené otázky a poznámky respondentov

Otázka	Texty
2	2-EURO,Oblasťsociológie-l'uds. vzťahy; 12-ohľadom dovolení; 18-ochutnáva jedál, potravinárskych výrobkov; 19-išlo o rôzne ochutnávky potravín; 22-kvalita produktov; 25-vyhodnotenie zozbieraných informácií a poznatkov prostredníctvom štatistických funkcií; 27-ste za zavedenie Eura; 33-výskum potravín; 34-kvalita produktov, vstup do EU, zavedenie EURA;
3	1-rozbor dát; 4-nič; 6-triedenie dát; 7-štatistický výskum, triedenie a porovnávanie údajov; 8-triedenie informácií na menšie celky; 9-štatistické spracovanie údajov; 10-štatistické spracovanie, hodnotenie, triedenie dát; 11-spracovanie údajov; 12-percentuálny výpočet výsledkov odpovedí týkajúcich sa určitej problematiky, al. iných dát; 13-nič zvláštne; 14-rozbor informácií a teda aj získavanie inf. prostred. štatistiky; 15-hodnotenie, zhodnotenie a porovnávanie určitých ukazovateľov za účelom vytvorenia konkrétneho prehľadu o ... ; 16-vyhľadávanie, zisťovanie, spracovanie, hodnotenie dát, informácií – ich využitie v danom súbore; 17-zozbieranie dát, z ktorých môžeme spraviť rôzne grafy a výsledky potom zhodnotiť; 18-z daných údajov sa dá vypočítať priemer, čo nám dáva určitú predstavu o údajoch, analyzovať napr. verejnú mienku, spracovanie výsledkov; 19-ide o analýzu- rozbor dát, pri ktorom sa využívajú poznatky štatistiky a jej výsledkom sú štatistické údaje ako napr. početnosť výskytu určitého javu atď.; 20-zhromažďovanie a následné „rozobratie“ dát pomocou rôznych štatistických funkcií a ich následné vyhodnotenie pomocou počítača a príslušných programov; 21-štatistické údaje, ktoré sa spracúvajú a vyhodnocujú; 22-zisťovanie rozličných skutočností vyplývajúcich z dostupných dát; 23-rozbor; 24-spracovávanie, vyhodnocovanie štatistických dát; 27-vytváranie štatistiky; 28-triedenie údajov, informácií do istého systému; 29-zatriedenie údajov do skupín; 30-asi nejaké zatriedovanie alebo rozbor údajov, s ktorými pracujeme; 31-spracovanie a vyhodnocovanie štatistických dát; 32-rozbor štatistických údajov; 33-zatriedenie a skúmanie dát štatistickými metódami; 34-spracovanie veľkého množstva dát a ich následné vyhodnotenie; 35-zisťovanie štatistických údajov zo súboru informácií, napr. najčastejšie sa vyskytujúce znaky, javy; 36-rozbor dát pomocou štatistiky; 37-analýza dát (čísel) pomocou štatistických funkcií;
4	2-už je zavedené;16-táto otázka je troška od veci – už aj moja stará mama si na Euro zvykla;
5	16-nie, ani ju nezháňam; 37-zatiaľ nie, ale určite sa na ňu pozriem;
6	4-moja súkromná vec; 11-Slobodné fórum; 15-nepriznám;
7	2-Monika Beňová; 3-Lipšic, Mikloš; 4-žiadny; 7-Kaliňák; 8-ani jednému; 9-Fico; 10-Fico; 11-Ivan Gašparovič; 12-Maďarič, Kaliňák, Fico; 13-nikomu; 14-Ficovi; 15-Fico, Lipšic, Kaliňák; 16-nikomu; 17-Kaliňák; 18-ťažko povedať; 19-Kaliňák;

	20-ani jednému; 21-Fico; 22-žiadnemu; 24-A. Malíková-Belousovová; 25-Dušan Čaplovič; 26-žiadnemu; 27-Fico; 28-žiadnemu; 30-Ficovi; 31-Robert Fico, Bella Bugár, Róbert Kaliňák; 32-Fico; 33-žiadnemu; 34-Robert Fico, Robert Kaliňák, Ivan Gašparovič; 35-Fico; 36-Radičová;
11	15-neviem, nestravujem sa v nej;
13	4-neviem posúdiť; 16-no comment; 17-očakávala som niečo viac, ale v zásade áno; 20-celkom áno;
14	15-nepáči sa mi polozenie otázky č.6, slová (príkaz) „priznajte sa“!; 16-za zvláštny; 20- (nechápem) neviem účel dotazníka; 27-nie;

Dotazník ilustruje aj niektoré nedostatky, ktoré si respondenti všimli.

Príkladom sú otázky:

č. 4, pri ktorej mali niektorí respondenti „problém“ s jej aktuálnosťou,

č. 5 – táto otázka uvádza, aj na pobavenie, neexistujúcu učebnicu – iste by sme získali zaujímavejšie odpovede, ak by sme dali napríklad študentom známeho autora – napr. doc. X.Y.,

č. 11 – nezohľadňuje respondentov, ktorí sa v jedálni nestravujú,

Doc. X.Y. zozbierala odpovede od 37 respondentov. Záznam dát v Exceli je v tabuľke. Otázky č. 2 a č.3 stačí vyhodnotiť „verbálne“.

Na príklade otázky č. 7 ilustrujeme kódovací kľúč, podľa ktorého boli zaznamenané odpovede.

Kódovací kľúč otázky č.7

- 1- Robert Fico
- 2- Róbert Kaliňák
- 3- Daniel Lipšic
- 4-Ivan Gašparovič
- 5-Monika Beňová;
- 6- Ivan Mikloš
- 7- Marek Maďarič
- 8- Dušan Čaplovič
- 9- Béla Bugár
- 10- Anna Malíková-Belousovová
- 11-Iveta Radičová
- 12-žiadnemu

Tabuľka na záznam dát

porcis	o1	o2a	o2b	o2c	o3a	o3b	o4	o5	o6	o7a	o7b	o7c	o8	o9	o10	o11	o12	o13	o14
1	3	.	.	.	.	.	5	5	11	.	.	.	2	22	1	2	3	1	2
2	2	.	.	.	.	.	5	5	10	5	.	.	1	22	1	5	2	1	3
3	3	.	.	.	.	.	1	5	4	3	6	.	1	22	1	.	2	4	3
4	3	.	.	.	.	.	1	5	11	12	.	.	1	23	1	3	3	3	4

5	3	.	.	.	.	.	2	5	11	.	.	.	1	22	1	4	4	1	4
6	3	.	.	.	.	.	1	5	10	.	.	.	2	22	1	.	3	1	3
7	3	.	.	.	.	.	1	5	6	2	.	.	2	23	1	3	1	1	2
8	3	.	.	.	.	.	1	5	11	12	.	.	1	23	2	3	2	1	2
9	3	.	.	.	.	.	4	5	6	1	.	.	2	22	1	3	2	1	3
10	3	.	.	.	.	.	1	5	6	1	.	.	2	23	1	3	2	1	3
11	3	.	.	.	.	.	2	5	9	4	.	.	1	22	1	4	2	1	3
12	2	.	.	.	.	.	1	5	6	7	2	1	2	22	1	3	2	1	3
13	3	.	.	.	.	.	4	5	10	12	.	.	1	22	1	3	4	1	4
14	3	.	.	.	.	.	3	4	6	1	.	.	2	21	1	1	3	1	1
15	3	.	.	.	.	.	1	5	.	1	3	2	1	23	1	.	2	1	5
16	3	.	.	.	.	.	4	6	11	12	.	.	2	23	1	2	3	3	5
17	3	.	.	.	.	.	4	5	10	2	.	.	2	23	1	4	3	3	2
18	2	.	.	.	.	.	2	5	11	12	.	.	2	22	1	3	2	1	2
19	2	.	.	.	.	.	3	5	6	2	.	.	2	21	1	2	2	1	2
20	3	.	.	.	.	.	2	5	10	12	.	.	1	22	1	4	2	3	5
21	3	.	.	.	.	.	2	5	10	1	.	.	1	24	1	.	1	1	1
22	2	.	.	.	.	.	2	5	11	12	.	.	1	24	1	2	3	1	3
23	3	.	.	.	.	.	1	5	4	.	.	.	1	.	1	2	2	1	2
24	3	.	.	.	.	.	4	5	5	10	.	.	2	22	1	.	3	1	2
25	3	.	.	.	.	.	1	5	4	8	.	.	1	23	1	2	3	1	1
26	3	.	.	.	.	.	3	5	11	12	.	.	1	22	1	4	2	1	2
27	2	.	.	.	.	.	1	6	6	1	.	.	1	23	1	2	2	1	5
28	3	.	.	.	.	.	2	5	11	12	.	.	2	22	1	3	2	1	2
29	3	.	.	.	.	.	3	5	2	.	.	.	1	22	1	.	3	1	2
30	3	.	.	.	.	.	3	5	6	1	.	.	2	22	1	3	2	1	2
31	3	.	.	.	.	.	2	5	4	1	9	2	2	22	1	3	3	1	1
32	3	.	.	.	.	.	1	5	6	1	.	.	2	21	1	2	2	1	1
33	2	.	.	.	.	.	5	5	10	12	.	.	1	22	1	5	3	1	3
34	2	.	.	.	.	.	1	5	6	1	2	4	1	23	1	2	3	1	2
35	3	.	.	.	.	.	2	5	10	1	.	.	1	23	1	2	2	1	1
36	3	.	.	.	.	.	1	5	4	11	.	.	2	23	1	2	2	1	2
37	3	.	.	.	.	.	1	6	11	.	.	.	2	24	1	3	1	1	1

Po zázname dát ich načítame do štatistického softvéru, prípadne môžeme mnohé výpočty urobiť v Exceli – návody sú napríklad v knihách J. Chajdiaka – presné odkazy sú v zozname literatúry.

Autor tejto práce uprednostňuje SPSS a tak si ukážeme dáta načítané do SPSS, kde môžeme zaznamenať aj texty otázok a odpovedí, čo je výhodné pri skúmaní výstupných zostáv.

8. Kontrola a čistenie dát

Dôležitým krokom pred analýzou dát je ich kontrola. Pred záznamom sme dotazníky skontrolovali, očíslovali. Kontrolu dát výhodne robíme napríklad v SPSS, kde máme možnosť zadať do súboru popisky dát a texty otázok (metadáta).

Prvý krok kontroly spočíva vo výpočte frekvenčných tabuliek. Tu uvedieme jeden príklad:

o1 Boli ste už oslovení agentúrou na výskum trhu a verejnej mienky?

		Frequenc y	Percent	Valid Percent	Cumulati ve Percent
Valid	2 áno viackrát	7	18,9	19,4	19,4
	3 nie	28	75,7	77,8	97,2
	6	1	2,7	2,8	100,0
	Total	36	97,3	100,0	
Missing	System	1	2,7		
Total		37	100,0		

V tabuľke je jeden chybný údaj, kód 6 a jeden chýbajúci údaj – ľahko môžeme identifikovať číslo dotazníka a chyby opraviť.

Okrem kontroly na rozsah, ktorú sme ilustrovali, máme možnosti ďalšej kontroly, ktorá spočíva v logickom kontrole, prípadne aj kontrole odľahlých dát.

Pri logickej kontrole môžeme využiť kontingenčné tabuľky – napríklad pri zistení, či boli dodržané filtre, alebo logika skúmaných otázok.

V našom dotazníku máme filter – teda na otázku číslo 2 neodpovedajú všetci respondenti. Využijeme kontingenčnú tabuľku o1 s o2. Vieme, že na o2 nemali odpovedať respondenti, ktorí v o1 odpovedali kódom 3. Pomocou kontingenčnej tabuľky zistíme, že jeden respondent odpovedal i keď nemal – ľahko ho identifikujeme a zistíme v čom je chyba a dáta opravíme.

o1 Boli ste už oslovení agentúrou na výskum trhu a verejnej mienky? * o2a Ak áno, čo bolo obsahom dotazníka?**Crosstabulation**

Count

		o2a Ak áno, čo bolo obsahom dotazníka?	Total
		1 o trhu	
o1 Boli ste už oslovení agentúrou na výskum trhu a verejnej mienky?	3 nie	1	1
Total		1	1

9. Príprava dát na štatistické spracovanie

Po kontrole a čistení dát, ich pripravujeme na štatistickú analýzu. Príprava dát na štatistické spracovanie znamená okrem primárnych dát pripraviť aj transformácie ako napríklad indexy, podiely, ale aj kategorizácia napríklad veku, BMI a i.

Na ilustráciu prípravy dát na štatistické spracovanie uvádzame kategorizáciu BMI a veku a výpočet podielu (rate) úspešných oplodnení vajíčok.

BMI:

```
IF (BMI <= 25) BMI_kat = 1 .
IF (BMI > 25 and BMI <= 30) BMI_kat = 2 .
IF (BMI > 30) BMI_kat = 3 .
VALUE LABELS BMI_kat
1 "normal"
2 "nadváha"
3 "obezita"
```

vek_kat:

```
IF (vek <= 30) vek_kat = 1 .
IF (vek > 30) vek_kat = 2 .
```

$$FR_perc = 100 * FR_Men / FR_cit$$

sub	porcis	BMI_kat	BMI	vek	vek_kat	FR_cit	FR_men	FR_perc
1	1	1	24,03	26	1	10	6	60,00
1	2	1	23,56	30	1	8	4	50,00
1	3	1	23,56	30	1	15	12	80,00
1	9	1	23,25	37	2	25	19	76,00
1	10	1	18,82	30	1	7	6	85,71
1	11	1	24,97	37	2	27	17	62,96
1	12	1	21,13	32	2	5	5	100,00
1	13	1	18,59	31	2	7	7	100,00
1	24	1	20,44	31	2	11	9	81,82
1	25	1	19,15	28	1	6	4	66,67
1	41	1	24,97	39	2	2	2	100,00
1	42	1	19,83	30	1	12	6	50,00
1	43	3	38,1	33	2	4	4	100,00
1	44	3	30,58	30	1	5	4	80,00
1	45	2	28,71	31	2	6	6	100,00
1	46	3	35,43	20	1	7	6	85,71
1	47	3	30,71	29	1	8	6	75,00

Príklady transformácií dát pre ďalšie štatistické spracovanie ilustrujú ako dáta pripravíme a tak získame možnosti pre ďalšie analýzy – napríklad komparácie podľa vekových kategórií, kategórií BMI a nové možnosti na spracovanie transformovanej premennej FR.

10. Praktické príklady

V kapitole 2 uvádzame 5 príkladov súborov medicínskych dát. Na ilustráciu ich záznamu uvádzame „výrezy“ databáz z dvoch príkladov, tak aby sme neporušili autorské práva. Kvôli úspore miesta uvádzame iba niekoľko záznamov a iba časť skúmaných znakov. Kvôli ľahšej identifikácii ponechávame číslovanie príkladov z kapitoly 2.

Príklad 3: Národný register vrodenných vývojových chýb.

ROK	por c	week day	POHLA VIE	TYZDE N_T1	POC_P LOD	HMOTNOST	DLZKA	STAV_ DIETA	vek_ matky	KB1_M ATKY	vek_ otca	MKCH10	VCH_ KOM BIN
1 996	1	6	1	40	1	4 000	50	4		302	29	Q54	0
1 996	2	7	1	40	1	3 100	48	4		802	47	Q170	0
1 996	3	5	1	38	1	2 600	48	4		710	27	Q543	0
1 996	4	2	1	38	1	3 850	50	4		606	28	Q07 Q03	1
1 996	5	4	1	41	1	3 600	55	4		803	46	Q660	0
1 996	6	4	1	40	1	3 800	51	4		305	31	Q74	0
1 996	7	2	2	38	1	2 700	48	2	28	801	31	Q791	0
1 996	8	7	2	36	1	2 100	38	1	45	703	45	Q000	0
1 996	9	5	1	40	1	2 300	46	4	45	202	49	Q90	0
1 996	10	3	2	39	1	2 240	46	4	44	108	27	Q90	0

Na doplnenie uvádzame aj vypočítané počty skupín VVCh za uvedené záznamy:

ROK	porc	Q00_ Q07	Q10_ Q18	Q20_ Q28	Q30_ Q34	Q35_ Q37	Q38_ Q45	Q50_ Q56	Q60_ Q64	Q65_ Q79	Q80_ Q89	Q90_ Q99	INE	VVCh_ Q	VVCh_ all
1 996	1	0	0	0	0	0	0	1	0	0	0	0	0	1	1
1 996	2	0	1	0	0	0	0	0	0	0	0	0	0	1	1
1 996	3	0	0	0	0	0	0	1	0	0	0	0	0	1	1
1 996	4	2	0	0	0	0	0	0	0	0	0	0	0	2	2
1 996	5	0	0	0	0	0	0	0	0	1	0	0	0	1	1
1 996	6	0	0	0	0	0	0	0	0	1	0	0	0	1	1
1 996	7	0	0	0	0	0	0	0	0	1	0	0	0	1	1
1 996	8	1	0	0	0	0	0	0	0	0	0	0	0	1	1
1 996	9	0	0	0	0	0	0	0	0	0	0	1	0	1	1
1 996	10	0	0	0	0	0	0	0	0	0	0	1	0	1	1

Príklad 4: Databáza Abortus.

Pôvodné kódovanie databázy Abortus stanovené v roku 1992 bolo nutné rekódovať podľa všeobecných zásad daných v kapitole 4. Textové kódy sú transformované na numerické a súčasná verzia databázy umožňuje maximálne jej využitie na štatistické analýzy. Pripravovaná je príručka kódov databázy. Na ilustráciu uvádzame zase iba výrez databázy. Kódovanie karyotypov ešte nie je ukončené a tak ilustračný príklad uvádza pôvodné kódy.

Kult_c	Morf_K	Morfológia1	Morfológia2	Morfológia3	Kar_K	Karyotyp	Trimester	vek_plodu	vek_matky	pohl
34	3					0	1	6		
35	6					0	1			
36	5	gastroschisis	mesenterium commune			1 46, XY	2	22	25	1
37	6					0	1			
38	6					0	1			
39	8					0	1	5		
91	5	VVC	Down sy			2 47,XY,+G	2		43	1
92	6					0	1			
95	10					0	1	5		
96	2					0	1			
98	9					1 46, XY	1	10		1
99	10					0	1	8		
100	1	MHP				0	1	6		

11. Štatistický softvér (Excel a profesionálne softvéry)

Profesionálny štatistický softvér poskytuje aj možnosti na záznam dát, ich kontrolu a čistenie. Na veľké rozsahy dát je výhodné realizovať záznam dát elektronicky (už sa tešme na elektronické zdravotníctvo).

MS Excel je vzhľadom na všeobecnú dostupnosť vhodným prostriedkom na záznam dát.

Stručne sa zmienime o profesionálnom štatistickom softvéri, voľnom a čiastočne voľnom štatistickom softvéri. Bližšie o štatistike v Exceli bude v iných prednáškach podľa programu MEDSTAT 2010.

Profesionálny štatistický softvér

SPSS, patrí medzi najrozsiahlejšie a najrozšírenejšie profesionálne štatistické softvéry. Vzhľadom na používateľsky priateľské ovládanie, vynikajúci data editor a najlepšie možnosti pre spracovanie multiresponse znakov autor tohto príspevku jednoznačne uprednostňuje tento softvér. Viac informácií na stránkach www.spss.com a www.spss.cz.

SAS, taktiež patrí medzi najrozsiahlejšie a najrozšírenejšie profesionálne štatistické softvéry. Viac informácií na www.sas.com.

STATISTICA, vynikajúci profesionálny štatistický softvér, ktorý má už aj českú lokalizáciu. Vynikajúca elektronická učebnica je voľne dostupná na stránke www.statsoft.com.

SYSTAT, výborný profesionálny štatistický softvér. Má free študentskú verziu **MYSTAT**, jeden z cenovo dostupnejších. Bol nejaký čas súčasťou rodiny SPSS, ale znovu je rozvíjaný samostatne – vid': <http://www.systat.com>.

NCSS, ďalší výborný štatistický produkt, <http://www.ncss.com>. Obsahuje moduly na výpočet rozsahu výberu a silu štatistických testov PASS a GESS špeciálny softvér na spracovávanie štatistiky génov.

GENSTAT, softvér určený pre biovedu, napríklad na environmentálne dáta, poľnohospodárske dáta, biologické dáta. <http://www.vsn.co.uk/software/genstat>.

WINKS, menší ale cenovo dostupný iba 99 USD za Basic verziu a 229 USD za Professional verziu (<http://www.texasoft.com>). Obsahuje značný rozsah štatistických metód a grafov. Čiastočne voľný, umožňuje realizovať časť analýz na redukovanej množine dát – do 100 záznamov a limitovaný počet znakov, neumožňuje dáta ukladať.

MEDCALC, čiastočne voľný – na 30 otvorení, menší štatistický softvér špeciálne venovaný medicínskym analýzám ROC kriviek. <http://www.medcalc.be/>

SigmaPlot, obsahuje aj SigmaStat: <http://www.aspiresoftwareintl.com/html/sigmaplot.html>.

Voľný (free) štatistický softvér

Na internete možno nájsť veľa voľného softvéru, na ilustráciu vyberieme dva.

EPIINFO, je určený na spracovanie epidemiologických dát, medicínskych a zdravotníckych dát. <http://www.cdc.gov/epiinfo>.

VISTA, špeciálny štatistický softvér na vizualizáciu výsledkov niektorých štatistických analýz. Možno ho získať na stránke: <http://forrest.psych.unc.edu/research>.

R – voľne šíriteľný programovací jazyk s bohatou knižnicou aplikácií vrátane štatistických. Programovací prostriedok možno získať napríklad na stránke <http://www.r-project.org/>.

OpenStat, voľne dostupný štatistický softvér, nájdeme na stránke <http://statpages.org/miller/openstat/>.

Ďalší free štatistický softvér poskytujú stránky: <http://en.freestatistics.info/stat.php>, http://statistiksoftware.com/free_software.html; <http://statpages.org/miller/openstat>.

Prakticky dostupná je štatistika v Exceli, keď máte MS Excel.

12. Literatúra

- [1] Böhmer, D., Brašeňová D., Luha, J.(2009): Špecifiká záznamu a spracovania dát z Národného registra pacientov s vrodenou vývojovou chybou. FORUM STATISTICUM SLOVACUM 7/2009. SŠDS Bratislava 2009. ISSN 1336-7420.
- [2] Chajdiak J.(2005): Štatistické úlohy a ich riešenie v Exceli. STATIS, Bratislava, ISBN 80-85659-39-5.
- [3] Chajdiak J.(2007): Štatistika v exceli 2007. Statis, Bratislava 2009, ISBN 978-80-85659-49-8.
- [4] Luha J.(2003a): Skúmanie súboru kvalitatívnych dát. EKOMSTAT 2003. SŠDS Bratislava 2003.

- [5] Luha J.(2003b): Matematickoštatistické aspekty spracovania dotazníkových výskumov. Štatistické metódy vo vedecko-výskumnej práci 2003, SŠDS, Bratislava 2003. ISBN 80- 88946-32-8.
- [6]Luha J.(2008a): Prvotná štatistická analýza kvalitatívnych dát. FORUM STATISTICUM SLOVACUM 2/2008. SŠDS Bratislava 2008. ISSN 1336-7420.
- [7]Luha J.(2008b): Metodologické aspekty zberu a záznamu dát otázok s možnosťou viac odpovedí. FORUM STATISTICUM SLOVACUM 7/2008. SŠDS Bratislava 2008. ISSN 1336-7420.
- [8]Luha J.(2010): Základné oblasti a úlohy medicínskej štatistiky. FORUM STATISTICUM SLOVACUM 1/2010. SŠDS Bratislava 2010. ISSN 1336-7420.
- [9]Pecáková I.(2008): Statistika v terénnych průzkumech. Proffessional Publishing, Praha 2008. ISBN 978-80-86946-74-0.
- [10] Řezanková A.(2007): Analýza dat z dotazníkových šetření. Proffessional Publishing, Praha 2007. ISBN 978-80-86946-49-8.
- [11] <http://www.vszp.sk/showdoc.do?docid=231>

13. Internetové zdroje

Okrem literatúry, ktorá sa zaoberá záznamom dát možno nájsť na internete mnohé zdroje programov, ktoré umožňujú záznam dát a ich štatistické spracovanie. V práci [7] uvádzame viac internetových zdrojov, ktoré informujú o štatistickom softvéri. Informácie o internetových zdrojoch pre štatistický softvér sú aj v kapitole 11. Nasledujúce zdroje rozširujú informácie o ďalšie zdroje s dôrazom na záznam dát.

http://spreadsheets.about.com/od/excel101/ss/enter_data.htm
http://www.medcalc.be/manual/data_checking.php
http://www.medcalc.be/manual/summary_statistics.php
<http://www.deakin.edu.au/~agoodman/sci101/index.php>
<http://people.umass.edu/evagold/excel.html>
<http://home.ubalt.edu/ntsbarsh/stat-data/Topics.htm>
<http://home.ubalt.edu/ntsbarsh/excel/excel.htm>
http://spreadsheets.about.com/od/excel101/ss/enter_data.htm
http://www.internet4classrooms.com/excel_enter_edit.htm
http://www.medcalc.be/manual/how_to_enter_data.php
<http://sunsite.univie.ac.at/ViSta/>
<http://en.freestatistics.info/stat.php>
http://statistiksoftware.com/free_software.html
<http://statpages.org/miller/openstat/>
<http://www.texasoft.com/>
<http://www.aspiresoftwareintl.com/html/sigmaplot.html>
<http://www.r-project.org/>

Adresa autora:

Ján Luha, RNDr., CSc.

Ústav lekárskej biológie, genetiky a klinickej genetiky LF UK a FNsP

Sasinkova 4, Bratislava

jan.luha@fmed.uniba.sk

Kontrola súborov vstupných diskretných hodnôt pomocou podsystému Pivot Table systému Excel verzie 2007 Verification file input discrete values by subsystem Pivot Table of the system Excel (version 2007)

Jozef Chajdiak

Abstract: Report includes description technique of the inspection file input discreet values, identification sector incorrectly values, assignment in those observation myself these incorrect values find and their correction. We use to solution subsystem Pivot Table of the system Excel (version 2007).

Abstrakt: Príspevok obsahuje opis postupu kontroly súborov vstupných diskretných hodnôt údajov, identifikácie časti nekorektných hodnôt, špecifikácie v ktorých pozorovaniach sa tieto chybné hodnoty nachádzajú a ich opravu. K riešeniu sa používa podsystém Pivot Table systému Excel verzie 2007.

Key words: Pivot Table, Excel, file, variable, discrete values, error, correction.

Kľúčové slová: Pivot Table, Excel, súbor, premenná, diskretné hodnoty, chyby, opravy.

1. Úvod

Máme súbor zistených údajov. Tento súbor v prvom stĺpci obsahuje identifikačnú premennú (vhodné je označenie stĺpca i, I, ID resp. PC a ako hodnoty treba použiť postupnosť prirodzených čísiel 1, 2, 3, ..., n), ktorá jednoznačne identifikuje pozorovania (štatistickú jednotku, pacienta, pokus, ...). Okrem iných údajov, súbor obsahuje aj premenné s diskretnými hodnotami (napríklad pohlavie, špecifikácia štádia procesu, špecifikácia verzie choroby, špecifikácia dňa v týždni, mesiaca v roku a pod.).

Proces prenosu zistených diskretných údajov do elektronickej verzie môže byť, a zvyčajne aj je, sprevádzaný prítomnosťou chybných alebo nekorektných zápisov, či zistení, ktoré môžu ovplyvniť výsledok riešenia príslušnej analytickej úlohy. Chyby vo východiskových údajoch môžu negatívne ovplyvniť následné riadiace rozhodnutie (napríklad rozhodnutie o chybnom postupe liečenia zodpovedajúce chybnej hodnote v súbore a nie správny postup liečenia zodpovedajúci skutočnej hodnote z objektívnej reality).

Prvou úlohou vlastnej analýzy zistených a do elektronickej verzie prenesených hodnôt (hárok systému Excel) je kontrola ich korektnosti. Ku kontrole nie príliš početného množstva diskretných hodnôt analyzovanej premennej môžeme použiť podsystém systému Excel – podsystém Pivot Table (Poznámka: V slovenskej verzii Excelu sa používa označenie *Kontingenčná tabuľka*. Vzhľadom k tomu, že pojem *Kontingenčná tabuľka* sa používa aj na označenie konkrétnej verzie tabuľky, ktorá tiež môže byť výstupom Pivot Table a je len malou časťou z potenciálnych výstupov Pivot Table, autor preferuje použitie anglického pojmu Pivot Table a dúfa v jeho zdomácnenie v slovenčine).

2. Príklad súboru s chybami

K ilustrácii postupu identifikácie chybných hodnôt použijeme excelovský súbor **MEDSUBOR2010chyby.xlsx**, ktorý obsahuje údaje o meraní tlaku, tepu a hmotnosti respondenta v určitom časovom okamihu. V stĺpci B sú uložené konkrétne hodnoty

premennej SITUÁCIA špecifikujúcej, že ide o údaje PRED liečením, PO liečení a MEDZI začiatkom a koncom liečenia. Prezentované sú dve typické chyby – chybný syntaktický zápis - namiesto PO je napísané POO čo predstavuje novú ďalšiu hodnotu hoci ide len o preklep so zápisom jedného „O“ navyše (riadok I=1280). Druhú typickú chybu predstavuje zápis hodnoty z vedľajšieho stĺpca – v riadku I=1281 je miesto hodnoty „PO“ uvedený rok „2006“ z vedľajšieho stĺpca. V stĺpci denkod sú jednotlivé dni týždňa kódované číslami 1 až 7 (1 – pondelok, 2 – utorok, ..., 7 – nedeľa). V riadku I=1275 máme hodnoty „8“. Keďže týždeň má len sedem dní, tak ide o chybný zápis. Je však otázkou, či správna hodnota je 7 alebo 1 alebo 2 (zdá sa pravdepodobnejšie, že 1 – pondelok). V tomto súbore máme špecifikáciu aj dátumu (premenné ROK, mes a den) a tak môžeme jednoznačne určiť správnu hodnotu denkod ako 1 (pondelok). Poslednou ukážkou je sporná hodnota 1601 v stĺpci rokmes v riadku I=1277. Premenná rokmes obsahuje kódy mesiacov, v ktorých sa realizovalo pozorovanie – je štvorciferná, pričom prvé dve cifry obsahujú posledné dvojčísle letopočtu a druhé dve cifry poradové číslo mesiaca v roku. Hodnota 601 znamená január 2006 a hodnota 1601 znamená január 2016. Keďže január 2016 bude od momentu písania tohto príspevku až za 6 rokov zdá sa, že právnou hodnotou je „601“.

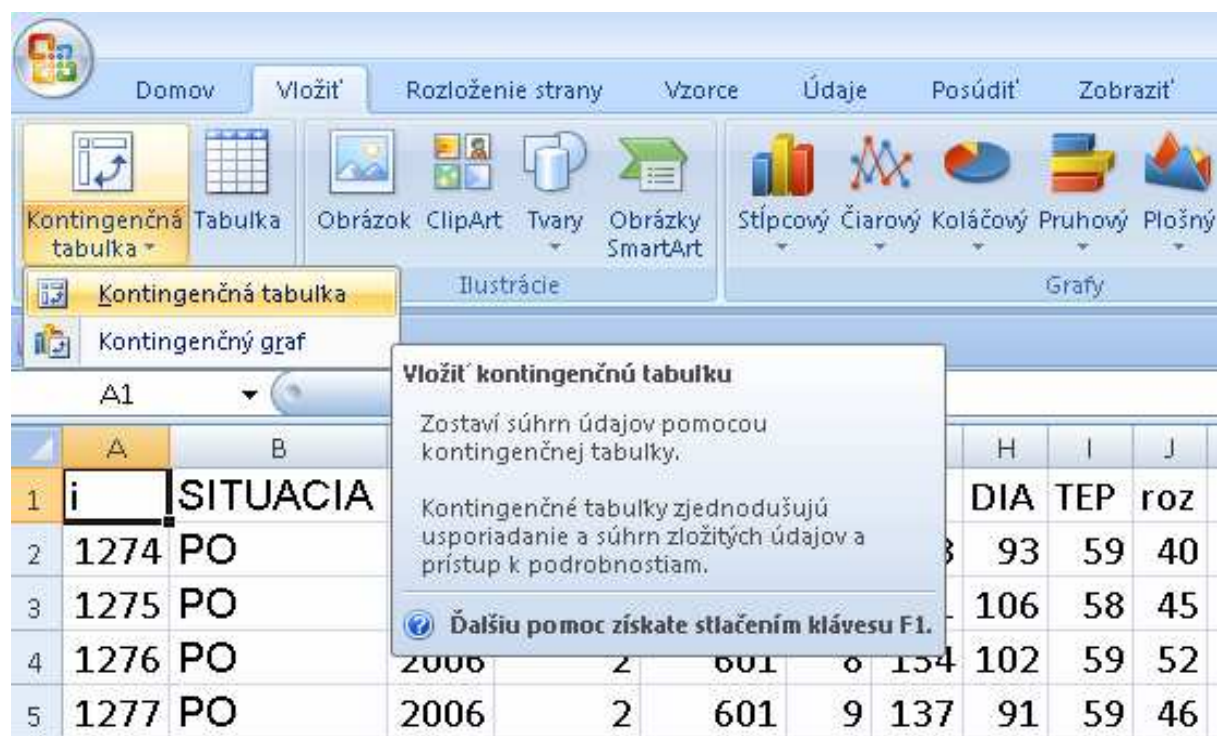
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	i	SITUACIA	ROK	denkod	rokmes	hod	SYS	DIA	TEP	roz	hmot	mes	den
2	1274	PO	2006	7	601	10	133	93	59	40	124,2	1	1
3	1275	PO	2006	8	601	9	151	106	58	45	123,9	1	2
4	1276	PO	2006	2	601	8	154	102	59	52	124,2	1	3
5	1277	PO	2006	2	1601	9	137	91	59	46	123,7	1	3
6	1278	PO	2006	3	601	8	154	102	66	52	124,3	1	4
7	1279	PO	2006	3	601	9	141	95	62	46	123,7	1	4
8	1280	POO	2006	4	601	8	155	102	64	53	123,1	1	5
9	1281	2006	2006	5	601	8	156	106	61	50	123,2	1	6
10	1282	PO	2006	6	601	8	151	98	58	53	123,8	1	7

Obrázok 1: Časť súboru MEDSUBOR2010chyby.xlsx

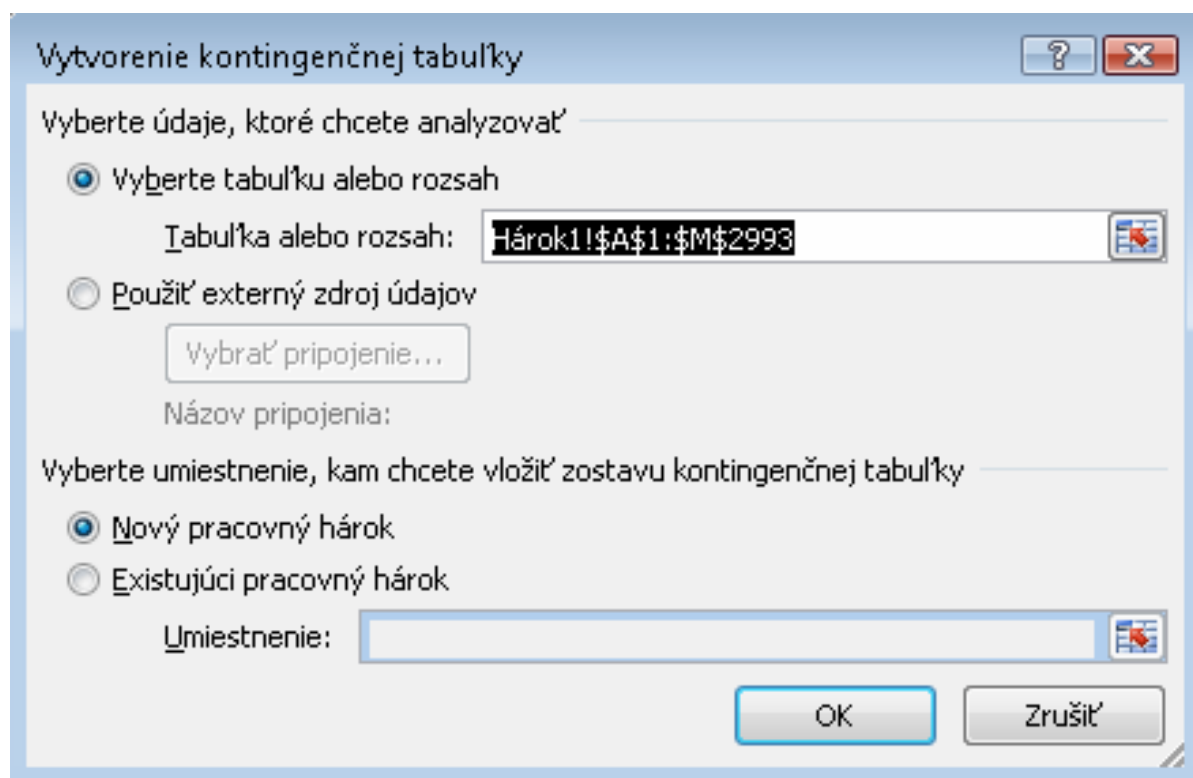
3. Identifikácia chybných hodnôt

K vlastnej identifikácii chybných alebo sporných hodnôt použijeme podsystém Pivot Table systému Excel (verzie 2007). Syntax použitia verzie 2007 je významne iný ako syntax starších verzií Excelu a nový Excel poskytuje aj ďalšie možnosti navyše.

Kurzor sa nachádza v poli tabuľky hodnôt v Hárku 1 otvoreného súboru MEDSUBOR2010chyby.xlsx. Podsystém Kontingenčná tabuľka (Pivot Table) aktivujeme postupnosťou príkazov Vložiť / Kontingenčná tabuľka / Kontingenčná tabuľka (Obrázok 2). Objaví sa automatická špecifikácia rozsahu analyzovaných údajov a miesta priebežného ukladania medzivýsledkov (Obrázok 3).



Obrázok 2: Aktivácia Kontingenčnej tabuľky (Pivot Table) – prvý krok



Obrázok 3: Aktivácia Kontingenčnej tabuľky (Pivot Table) – druhý krok

MEDSUBOR2010chyby

	A	B	C	D
1				
2				
3	Menovky riadkov	Počet z i		
4	2006	1		
5	PO	1717		
6	MEDZI	938		
7	POO	1		
8	PRED	335		
9	Celkový súčet	2992		
10				
11				
12				
13				
14				
15				
16				
17				
18				

Polia kontingenčnej tabuľky

Vyberte polia, ktoré chcete pridať do zostavy:

- ☒ i
- ☒ **SITUACIA**
- ☐ ROK
- ☐ denkod
- ☐ rokmes
- ☐ hod
- ☐ SYS
- ☐ DIA
- ☐ TEP
- ☐ roz

Presúvať polia medzi nižšie uvedenými oblastami:

Filter zostavy Menovky stĺpcov

Menovky riadkov Hodnoty

SITUACIA **Počet z i**

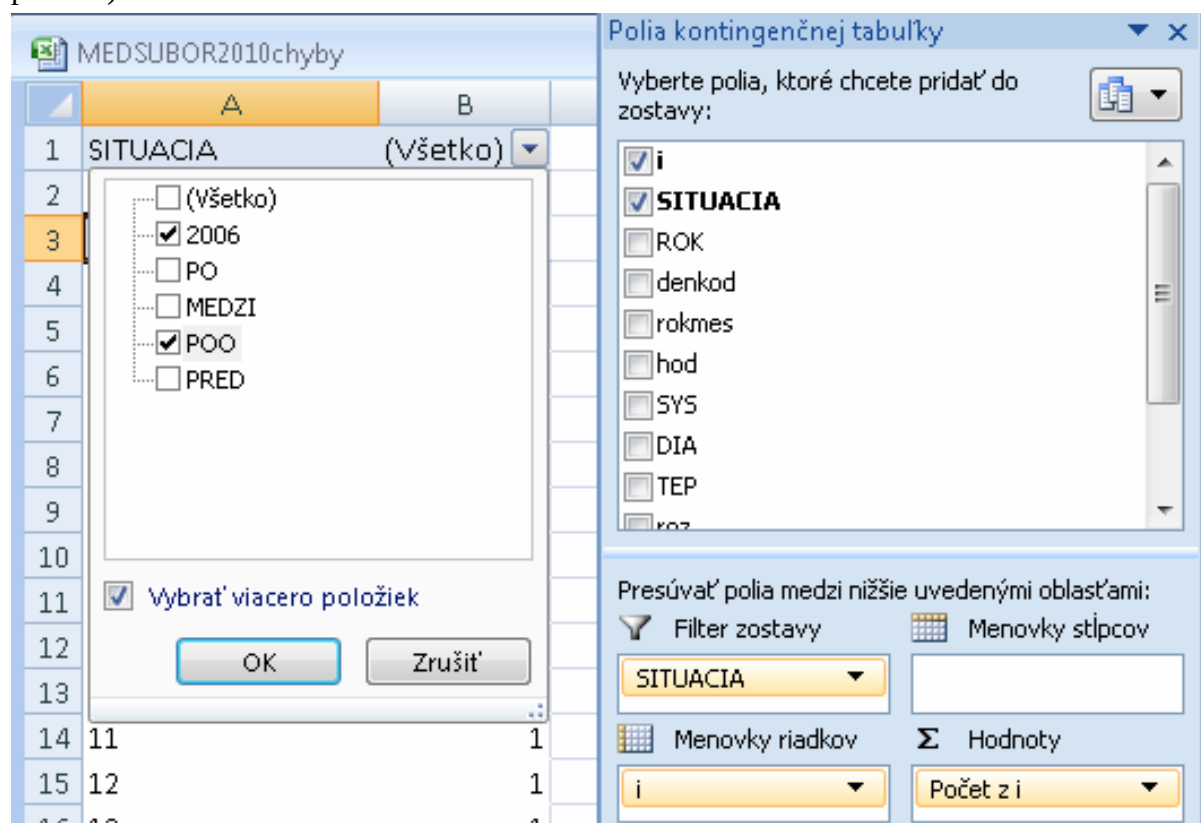
Obrázok 4: Špecifikácia jednotlivých hodnôt premennej SITUACIA

Vytriedime hodnoty premennej SITUACIA a tým skontrolujeme, či v súbore sú len prípustné hodnoty (PRED, MEDZI a PO) alebo aj iné hodnoty. Z časti so zoznamom premenných (vpravo hore na obrázku 4) presunieme do poľa Hodnoty (vpravo dole) premennú „i“ a zmeníme jej nastavenie z verzie „Súčet“ na verziu „Počet“. Do poľa Menovky riadkov presunieme premennú SITUACIA. Výsledok triedenia je v ľavej časti obrázku 4. Hodnota 2006 sa vyskytuje 1 krát, hodnota PO 1717 krát, hodnota MEDZI 938 krát, hodnota POO 1 krát a hodnota PRED 335 krát. Je zrejmé, že hodnoty 2006 a POO predstavujú chybné zápisy a je potrebné ich nahradiť správnou hodnotou „PO“. Keďže súbor obsahuje 2992 riadkov (pozorovaní) „ručná“ identifikácia riadku s hodnotou „POO“ alebo „2006“ môže byť dosť prácna. Použitie nástroja Pivot Table úlohu významne uľahčuje. Zápis riešenia je na obrázku 5 a výstup riešenia na obrázku 6.

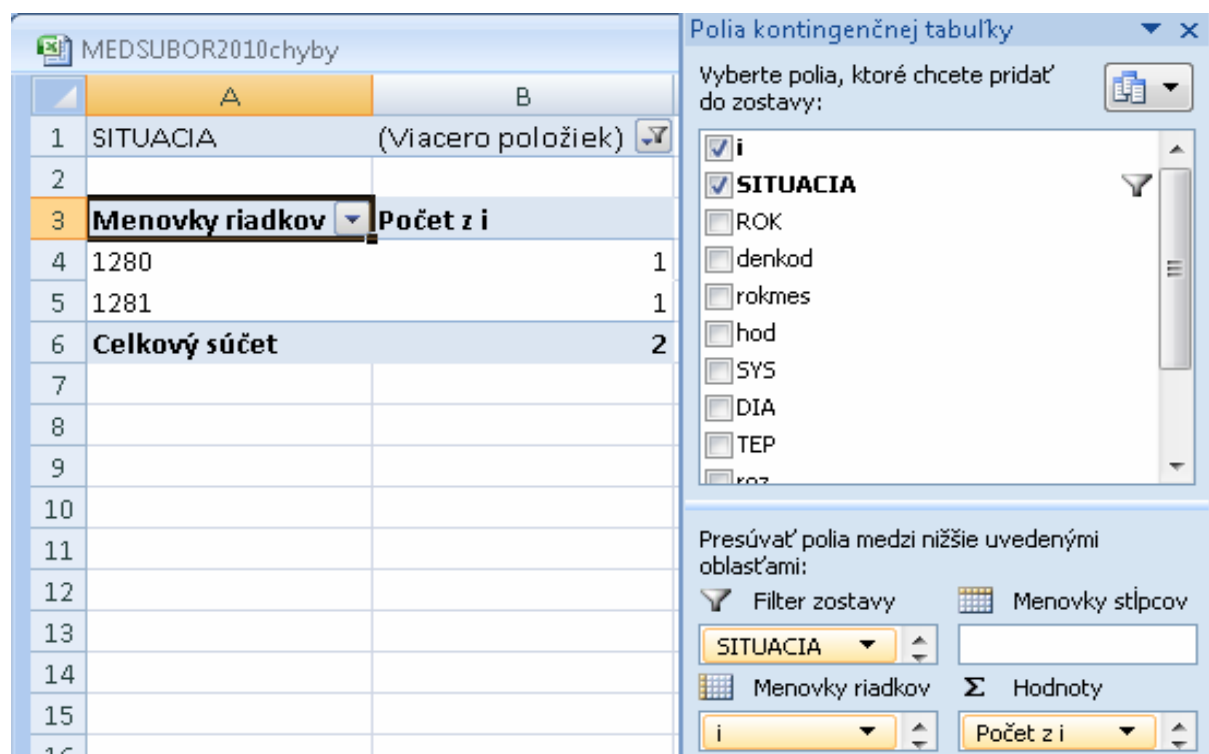
Z časti so zoznamom premenných (vpravo hore) presunieme do poľa Hodnoty (vpravo dole) premennú „i“ a zmeníme jej nastavenie z verzie „Súčet“ na verziu „Počet“. Do poľa Menovky riadkov presunieme opäť premennú „i“. Do poľa Filter zostavy presunieme premennú „SITUACIA“, aktivujeme filter a vyberieme hodnoty „2006“ a „PO“ (obrázok 5). Výstup je na obrázku 6. Chybné hodnoty sú v riadkoch I=1280 a I=1281. V Hárak1 ich opravíme na hodnotu „PO“ a vrátime sa kontrolovať ďalšie hodnoty.

Na obrázku 7 sú výsledky triedenia súboru podľa hodnôt premennej denkod. Máme jednu hodnotu navyše – kód „8“, ktorý je chybnou hodnotou a pred rozhodnutím, čím ho nahradiť musíme špecifikovať jeho miesto výskytu v súbore (obrázok 8). Je to riadok I=1275 a z analýzy ostatných hodnôt vyplýva, že hodnotu „8“ nahradíme hodnotou „1“. V prípade, že nedokážeme jednoznačne určiť akou hodnotou chybnú a nemožnú hodnotu „8“ v súbore nahradiť, pokúsime sa znovu zistiť príslušnú hodnotu danej premennej a ak je toto nové

zistenie nemožné, nemožnú hodnotu „8“ vylúčime bez náhrady zo súboru (ostane prázdne políčko).



Obrázok 5: Špecifikácia jednotlivých chybných hodnôt premennej SITUACIA



Obrázok 6: Identifikácia riadkov s chybnými hodnotami

Polia kontingenčnej tabuľky

Vyberte polia, ktoré chcete pridať do zostavy:

- ☒ i
- ☐ SITUACIA
- ☐ ROK
- ☒ denkod
- ☐ rokmes
- ☐ hod
- ☐ SYS
- ☐ DIA
- ☐ TEP
- ☐ rok

Presúvať polia medzi nižšie uvedenými oblastami:

Filter zostavy:

Menovky stĺpcov:

Menovky riadkov:

Hodnoty:

denkod ▼

Počet z i ▼

Obrázok 7: Špecifikácia jednotlivých hodnôt premennej denkod

Polia kontingenčnej tabuľky

Vyberte polia, ktoré chcete pridať do zostavy:

- ☒ i
- ☐ SITUACIA
- ☐ ROK
- ☒ denkod
- ☐ rokmes
- ☐ hod
- ☐ SYS
- ☐ DIA
- ☐ TEP
- ☐ rok

Presúvať polia medzi nižšie uvedenými oblastami:

Filter zostavy:

Menovky stĺpcov:

Menovky riadkov:

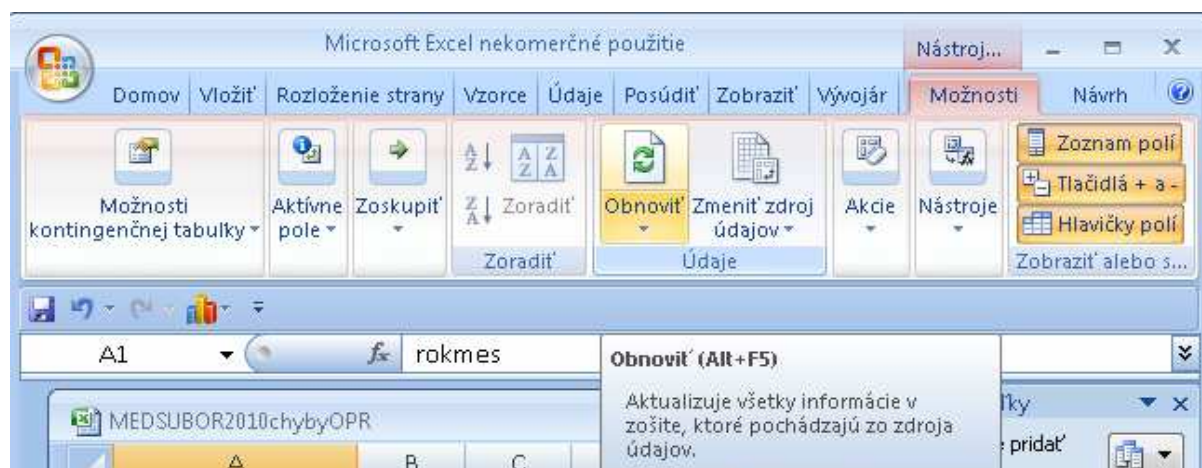
Hodnoty:

denkod ▼

i ▼

Počet z i ▼

Obrázok 8: Identifikácia riadku s chybnou hodnotou



Obrázok 9: Aktualizácia opraveného súboru

Po identifikácii chybných hodnôt a ich oprave v Hárak1 na korektné hodnoty je potrebné aktivovať prácu so súborom nových upravených hodnôt. Kurzor máme v poli tabuľky s priebežnými výsledkami podsystému Pivot Table a postupnosťou príkazov Možnosti / Nástroj / Obnoviť zmeníme starú verziu súboru na novú verziu.

4. Záver

- Úvodná práca s údajmi predstavuje snád' aj dve tretiny analytickej práce.
- Namýšľať si, že naša práca je bezchybná je ilúzia.
- Zdá sa výhodnejšie skontrolovať údaje na začiatku práce, ako ich kontrolovať po prácnom pátraní po príčinách nejakého „veľmi nečakaného“ výsledku .

5. Literatúra

- [1] CHAJDIK, J. 2003. Štatistika jednoducho. Bratislava: Statis 2003. 194 s. ISBN 80-85659-28-X.
- [2] Chajdiak J. (2009): Štatistika v Exceli 2007. STATIS, Bratislava, ISBN 80-85659-49-8.
- [3] Luha, J. (2006): *Štatistické metódy analýzy kvalitatívnych znakov*. FORUM STATISTICUM SLOVACUM 2/2006. SŠDS Bratislava 2006. ISSN 1336-7420.
- [4] Luha J.: Prvotná štatistická analýza kvalitatívnych dát. FORUM STATISTICUM SLOVACUM 2/2008. SŠDS Bratislava 2008. ISSN 1336-7420.
- [5] Luha J.(2010): Metodologické zásady záznamu dát z rozličných oblastí medicíny a zásady ich kontroly. FORUM STATISTICUM SLOVACUM 1/2010. SŠDS Bratislava 2010. ISSN 1336-7420.

Adresa autora:

Jozef Chajdiak, Doc., Ing., CSc.

Ústav manažmentu STU

Bratislava

Jozef.Chajdiak@stuba.sk

chajdiak@statis.biz

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

Súbory údajov pre MEDSTAT2010 Input file for MEDSTAT 2010

Jozef Chajdiak

Abstract: Report includes description five files for the work during MEDSTAT 2010.

Key words:, file, variables.

Kľúčové slová: súbor, premenné.

1. Úvod

Na webovskej stránke **www.statis.sk/medstat** sú k dispozícii vybrané súbory pre samostatnú prácu účastníkov konferencie MedStat, ako aj iných záujemcov. Je to aj 5 súborov vo verzii excelu verzie 1997-2003 (prípona xls) resp. excelu verzie 2007 (prípona xlsx). Konkrétne mená súborov sú nasledujúce:

TEP_2010.xls
TEP_2010.xlsx
ANALYZA_TEP_2010.xls
ANALYZA_TEP_2010.xlsx
TEP_TLAK_2010.xls
TEP_TLAK_2010.xlsx
MEDSUBOR2010.xls
MEDSUBOR2010.xlsx
MEDSUBOR2010chyby.xls
MEDSUBOR2010chyby.xlsx

Dvojice mien predstavujú obsahovo zhodné súbory, len jeden je uchovaný s staršej verzii excelu (prípona xls) a druhý je v novšej verzii excelu (prípona xlsx).

Súbory možno aktivovať príkazom Internet Exploreru, napríklad:

http://www.statis.sk/medstat/TEP_2010.xls

či

http://www.statis.sk/medstat/TEP_2010.xlsx

2. Súbor TEP_2010

Súbor TEP_2010 obsahuje 1719 pozorovaní o tepe (pulzovej frekvencii srdca). Súbor predstavuje podmnožinu súboru MEDSUBOR2010.xls pre pozorovania s hodnotou premennej SITUACIA = PO. Má dve premenné:

ID - identifikačná premenná pre každé pozorovanie (obsahuje postupnosť prirodzených čísiel od 1 po 1719)

TEP - tep (pulzová frekvencia srdca).

3. Súbor ANALYZA_TEP_2010

Súbor ANALYZA_TEP_2010 obsahuje opäť 1719 pozorovaní o tepe respondentov a jeho štatistickej analýze spracovanej v piatich hárkoch (tep1, tep2, ..., tep5).

Hárok **tep1** obsahuje východiskové hodnoty v premenných **i** a TEP:

i - identifikačná premenná pre každé pozorovanie (obsahuje postupnosť prirodzených čísiel od 1 po 1719)

TEP - tep (pulzová frekvencia srdca).

Na konci súboru je vypočítaný počet nechýbajúcich numerických hodnôt, minimálna a maximálna hodnota tepu.

Hárok **tep2** obsahuje hodnoty háрку **tep1** a navyše horné hranice tried (v políčkach D2 až D14).

Hárok **tep3** obsahuje obsah háрку **tep2** a absolútne triedne početnosti - výsledky funkcie FREQUENCY (políčka E2 až E15). Výsledky triedenia sú prezentované v úplnej frekvenčnej tabuľke a stĺpcovom grafe (histograme)

Hárok **tep4** obsahuje výstupy analytického nástroja Histogram – časti frekvenčnej tabuľky s absolútnymi a kumulatívnymi percentuálnymi početnosťami v klasickej verzii usporiadania a v Paretovskej verzii usporiadania riadkov tabuľky. Hárok obsahuje aj stĺpcový graf (histogram) vychádzajúci z Paretovskej frekvenčnej tabuľky doplnený o krivku percentuálnych početností.

Hárok **tep5** obsahuje hodnoty háрку **tep1**. Od riadku 1722 sú doplnené výsledky základného štatistického rozboru premennej TEP, úhrn, percentily, kvartily a poradia zodpovedajúce hodnotám (inverzná funkcia k percentilom).

4. Súbor TEP_TLAK_2010

Súbor TEP_TLAK_2010 obsahuje hodnoty za 2992 pozorovaní v desiatich premenných, oknkkrétne:

i - poradové číslo pozorovania (1, 2, ..., 2992),

SITUACIA – dosahuje tri hodnoty – PRED, MEDZI a PO, ktorým zodpovedajú pozorovania realizované PRED prijatím systému opatrení na zlepšenie zdravotného stavu, pozorovania PO zakončení systému opatrení na zlepšenie zdravotného a pozorovania MEDZI začiatkom a koncom fungovania systému opatrení na zlepšenie zdravotného stavu,

ROK – letopočet, v ktorom bolo pozorovanie realizované,

denkod – kód dňa (1 – pondelok, 2 – utorok, ..., 7 – nedeľa),

rokmes – posledné dvojčísle roku a dvojčísle vyjadrujúce mesiac (1003 – marec roku 2010),

hod – hodina merania parametrov,

SYS – systolický tlak,

DIA – diastolický tlak,

TEP – tep - pulzová frekvencia srdca

roz – rozdiel = SYS – DIA

5. Súbor MEDSUBOR2010

Súbor **MEDSUBOR2010** obsahuje hodnoty za 2992 pozorovaní v trinástich premenných, konkrétne:

i - poradové číslo pozorovania (1, 2, ..., 2992),

SITUACIA – dosahuje tri hodnoty – PRED, MEDZI a PO, ktorým zodpovedajú pozorovania realizované PRED prijatím systému opatrení na zlepšenie zdravotného stavu, pozorovania PO zakončení systému opatrení na zlepšenie zdravotného a pozorovania MEDZI začiatkom a koncom fungovania systému opatrení na zlepšenie zdravotného stavu,

ROK – letopočet, v ktorom bolo pozorovanie realizované,

denkod – kód dňa (1 – pondelok, 2 – utorok, ..., 7 – nedeľa),

rokmes – posledné dvojčísle roku a dvojčísle vyjadrujúce mesiac (1003 – marec roku 2010),

hod – hodina merania parametrov,

SYS – systolický tlak

DIA – diastolický tlak,

TEP – tep - pulzová frekvencia srdca

roz – rozdiel = $SYS - DIA$,

hmot – hmotnosť respondenta (v kg),

mes – mesiac pozorovania (1 –január, 2 – február, ..., 12 – december),

den – deň pozorovania (poradové číslo dňa v mesiaci).

6. Súbor MEDSUBOR2010chyby

Súbor **MEDSUBOR2010chyby** obsahuje, okrem bezchybných aj chybné hodnoty, za 2992 pozorovaní v trinástich premenných, konkrétne:

i – poradové číslo pozorovania (1, 2, ..., 2992),

SITUACIA – dosahuje tri hodnoty – PRED, MEDZI a PO, ktorým zodpovedajú pozorovania realizované PRED prijatím systému opatrení na zlepšenie zdravotného stavu, pozorovania PO zakončení systému opatrení na zlepšenie zdravotného a pozorovania MEDZI začiatkom a koncom fungovania systému opatrení na zlepšenie zdravotného stavu,

ROK – letopočet, v ktorom bolo pozorovanie realizované,

denkod – kód dňa (1 – pondelok, 2 – utorok, ..., 7 – nedeľa),

rokmes – posledné dvojčísle roku a dvojčísle vyjadrujúce mesiac (1003 – marec roku 2010),

hod – hodina merania parametrov,

SYS – systolický tlak

DIA – diastolický tlak,

TEP – tep - pulzová frekvencia srdca

roz – rozdiel tlakov = $SYS - DIA$,

hmot – hmotnosť respondenta (v kg),

mes – mesiac pozorovania (1 –január, 2 – február, ..., 12 – december),

den – deň pozorovania (poradové číslo dňa v mesiaci).

7. Literatúra

[1] Chajdiak J. (2009): Kontrola súborov vstupných diskretných hodnôt pomocou podsystému Pivot Table systému Excel verzie 2007. FORUM STATISTICUM SLOVACUM 1/2010. SŠDS Bratislava 2010. ISSN 1336-7420.

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

[3] Luha, J. (2006): *Štatistické metódy analýzy kvalitatívnych znakov*. FORUM STATISTICUM SLOVACUM 2/2006. SŠDS Bratislava 2006. ISSN 1336-7420.

[4]Luha J.: Prvotná štatistická analýza kvalitatívnych dát. FORUM STATISTICUM SLOVACUM 2/2008. SŠDS Bratislava 2008. ISSN 1336-7420.

[5]Luha J.(2010): Metodologické zásady záznamu dát z rozličných oblastí medicíny a zásady ich kontroly. FORUM STATISTICUM SLOVACUM 1/2010. SŠDS Bratislava 2010. ISSN 1336-7420.

Adresa autora:

Jozef Chajdiak, Doc., Ing., CSc.

Ústav manažmentu STU Bratislava

chajdiak@statis.biz

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

Neurónové siete v ekológii Neural networks in ecology

Mária Letkovičová a Hana Zach

Abstract:

The work studied the connection between the quality of groundwater and soil, and malignant breast tumors mortality. This is a transferred effect of nutrition, ingestion of contaminated water and food, or moderately defined, the imbalance of geochemical elements contained in food and water. Inputs were database of 34 environmental (geological and chemical) water indicators and 33 similar soil indicators, calculated as an average measurement for each municipality of the SR and the relative mortality from breast cancer in each municipality, expressed as a robust data for 22 years (1986 to 2008). Mathematical tools are neural networks. Best of 200 networks identified as powerful 12 indicators of water and soil, 55 environmental indicators have appeared as uninteresting in this context. It is considered that the impact of the surrounding environment to a specific problem, such as breast cancer, is now unrecognized and is certainly higher than is commonly assumed. At least it is not negligible and it is necessary that the experts began to address it with due seriousness.

Key words : Neural networks, breast cancer, ground water, soil, demografia

Kľúčové slová: Neurónové siete, rakovina prsníka, podzemná voda, pôda, demografia

1. Úvod

V tejto práci sa skúmal vzťah medzi kvalitou podzemnej vody a pôdy a úmrtnosťou na zhubné nádory prsníka. Ide o prenesený vplyv výživy, o prijímanie kontaminovanej vody a potravín, alebo miernejšie formulované, o nevyváženosť v potrave a vode obsiahnutých geochemických prvkov. Vstupnými údajmi bola databáza 34 environmentálnych (geologicko-chemických) indikátorov vody a 33 podobných indikátorov pôdy, vypočítaných ako priemerné meranie pre každú obec SR a relatívna úmrtnosť na rakovinu prsníka v každej obci, vyjadrenej ako robustný údaj za 22 rokov (1986 až 2008). Matematickým aparátom boli neurónové siete [1]. Najlepšia z 200 sietí určila ako vplyvné 12 indikátorov z vody a pôdy, 55 environmentálnych indikátorov sa javilo ako nezaujímavých v tejto súvislosti. Najvplyvnejším indikátorom bolo pH_{KCl} v pôde, ktoré sa očakáva nižšie, rozhodne pod hodnotou neutrality, ďalej bárium vo vode, ktoré by vôbec nemalo byť, atď. Niektoré indikátory boli jednoznačne negatívne (čím menej, tým lepšie), tri však mali podobu paraboly, kde bolo optimálne dosiahnuť určitú hladinu. Ako protektívne sa javí určité definované množstvo ceru, nátria a kobaltu v pôde. Pre vplyvné indikátory boli navrhnuté limitné hodnoty vzhľadom k najnižšej mortalite. Korelačný koeficient siete bol pomerne nízky. Spätný model hodnoty úmrtnosti na rakovinu prsníka vo všetkých obciach, vypočítaný iba z hodnôt vybraných 12 indikátorov, nebol vždy celkom presný. Ťažko sa modelovali obce, ktoré mali počet trvalých obyvateľov do 100, kde model pomerne vysoko presahoval, skutočné hodnoty boli výrazne nižšie, pretože boli často nulové a tiež nad 50.000 trvalých obyvateľov, kde bol model zase podhodnotený, skutočné hodnoty boli mierne vyššie. Pre obce v rozsahu 500 až 50.000 obyvateľov (spolu 2751 obcí) bol model vcelku prijateľný, priemerná odchýlka modelu od reality bola 2,07%. Z toho usudzujeme, že vplyv okolitého životného prostredia aj na taký špecifický problém, akým je rakovina prsníka, je dnes nedocenený a je rozhodne vyšší ako sa všeobecne predpokladá. Prinajmenšom nie je zanedbateľný a je potrebné, aby sa mu experti začali venovať s náležitou vážnosťou.

2. Teórie, dáta a metódy

Prípustnú mieru znečistenia životného prostredia určujú medzné hodnoty stanovené osobitnými predpismi. Tieto hodnoty sa stanovujú v súlade s dosiahnutým stavom poznania tak, aby sa neohrozovalo zdravie ľudí a aby sa neohrozovali ďalšie živé organizmy a ostatné zložky životného prostredia. Prekročením týchto limitovaných medzných hodnôt obsahu vzniká možnosť ohrozenia životného prostredia – *environmentálne riziko*. V prípade možností poškodenia, choroby alebo smrti človeka ako dôsledku pôsobenia rizikových faktorov vyskytujúcich sa v životnom prostredí hovoríme o *zdravotnom riziku*. Možnosť ohrozenia ostatných biotických zložiek životného prostredia je definovaná ako *ekologické riziko*. Predstavuje pravdepodobnosť výskytu alebo možnosť výskytu nepriaznivých ekologických účinkov ako výsledku vystavenia živých systémov jednému alebo viacerým stresorom (*metodický pokyn MŽP SR č. 632/98*). [2]

Ani pri podrobnom skúmaní sme nenašli v publikovaných materiáloch WHO žiadne špecializované výskumy na konkrétne rozbery zloženia pôd a podzemných vôd v spojitosti s environmentálnym zdravím. Na druhej strane však existuje vo svetovej literatúre nespočetné množstvo čiastkových a regionálnych prác pojednávajúcich o zhoršenom stave obyvateľstva z miestnej kontaminácie vôd, pôd a ovzdušia rôznymi chemickými prvkami. Tieto práce však vychádzajú len z konkrétnych lokálnych incidencií na rôzne ochorenia. Nemáme preto možnosť oprieť sa o existujúce medzinárodné zdravotné odporúčania. Skôr tvoríme základ do debaty pre podobné výstupy. Všeobecne sa uvádza, že zdravie je determinované z 50% genetikou, osobným prístupom konkrétnych ľudí a úrovňou zdravotnej starostlivosti, 30% patrí k socio ekonomickým záležitostiam a iba 20% ovplyvňuje konkrétne životné prostredie. Jednak sme nenašli konkrétneho autora tohto globálne známeho podielu, jednak si myslíme po preskúmaní našich dát, že by sa dalo o tomto zložení vplyvov s úspechom polemizovať. Domnievame sa, že podiel environmentu je všeobecne podceňovaný lebo sa o ňom príliš nevie a že je v skutočnosti oveľa vyšší.

2.1. Environmentálne indikátory

Pod environmentálnymi indikátormi rozumieme obsahy chemických prvkov v geologickom prostredí (vody, pôdy, sedimenty a horniny) vytvárajúce geochemické pozadie krajiny, ktoré či už pozitívne alebo negatívne vplyva na biotu a človeka. Obsahy chemických prvkov môžu byť vo vzťahu k biote a ľudskému zdraviu aj v deficite alebo nadbytku. Oba prípady je možné u určitej skupiny prvkov považovať za nepriaznivé.

Ako environmentálne indikátory v danej práci hodnotíme obsahy chemických prvkov - zložiek -parametrov v podzemných vodách a pôdach. Tieto dve geologické prostredia majú určite najväčší súvis s ľudským zdravím. Podzemné vody sú potenciálnym zdrojom pitných vôd a časť obyvateľstva Slovenskej republiky (20–30 %) využíva lokálne domáce studne a lokálne vodné zdroje na zásobovanie pitnou vodou. Pôda je na jednej strane základom potravinového reťazca a na strane druhej najmä detská populácia prostredníctvom náhodnej ingescie pôd je s ňou v bezprostrednom kontakte.

Súbor hodnotených regionálnych environmentálnych indikátorov je nasledovný:

pre podzemné vody: pH, MIN., ChSK_{Mn}, Ca+Mg, Li, Na, K, Ca, Mg, Sr, Fe, Mn, NH₄, F, Cl, SO₄, NO₂, NO₃, PO₄, HCO₃, SiO₂, Cr, Cu, Zn, As, Cd, Se, Pb, Hg, Ba, Al, Sb,

pre pôdy: Al, As, B, Ba, Be, Bi, Ca, Cd, Ce, Co, Cr, Cu, F, Fe, Hg, K, Mg, Mn, Mo, Na, Ni, P, Pb, Sb, Se, Sn, Sr, V, W, Zn, pH_{H2O}, pH_{KCl}, karbonáty.

Podkladom pre zostavenie environmentálnych indikátorov Slovenskej republiky bolo environmentálne - geochemické mapovanie prebiehajúce od začiatku 90. rokov 20. storočia a pokračujúce až do súčasnosti [3].

2.2. Príčiny vzniku zhubných nádorov prsníka

Rakovina prsníka je veľmi publikovaným momentom. Ak sa dnes (1.2.2010) zadá do *googlu* heslo „causes of breast cancer“, alebo po slovensky „príčiny rakoviny prsníka“, tak dostaneme po anglicky 16 900 000 odpovedí a po slovensky 4 870. Je jasné, že všetky zdroje nikdy nikto nepreskúma. Jedna z dobrých možností je tzv. meta analýza, t.j. zoradenie čo najväčšieho počtu literárnych odkazov podľa podobnosti obsahu a vyhodnotenie častosti jednotlivých názorov. Je možné, že to, čo sa tvrdí najčastejšie, bude bližšie k pravde. Alebo aspoň zistíme, čo sa prevažne hovorí a o čom sa nehovorí vôbec. Vybrali sme 10 najčastejšie citovaných prehľadových zdrojov, ktoré sú buď výsledkom práce veľkej kliniky (Mayo Clinic, USA), alebo prehľadovým dielom mnohých autorov, či oficiálnych veľkých ústavov [4]. Hneď prvý zdroj hovorí o tom, že je veľa rizikových momentov pre vznik rakoviny prsníka a pritom niektoré ženy ich majú všetky a rakovinu nedostali a niektoré nemajú ani jeden, iba tú rakovinu. Tiež sa často spomína, že rakovinu prsníka nikdy nespôsobí jediný samotný rizikový faktor, ale musí ich pôsobiť vždy viac súčasne. Celkovo sa nám podarilo zhrnúť podozrivé príčiny do 43 rizikových faktorov, z ktorých je iba 6 takých, kde sa uvažuje o nejakom špecifickom vplyve životného prostredia. Všetky ostatné vychádzajú iba z osobných pomerov postihnutej osoby. Nie je ani jediný faktor, ktorý by citovali všetky vybraté práce. Najčastejšie citované vplyvy sú genetické predpoklady, narastajúci vek, nadváha a strava vôbec, pravidelné fajčenie a pitie alkoholu, rakovina v rodine, bezdetnosť alebo neskorý prvý pôrod a liečba alebo diagnostika pomocou ionizujúceho žiarenia, najmä vo veku do 30 rokov. Veľmi podozrivé sú aj akékoľvek umelo dodávané hormóny týkajúce sa reprodukcie, počnúc antikoncepciou, končiac post menopauzálnou terapiou. Kuriózne sú napríklad antiperspiranty alebo implantáty, kde jeden súbor názorov ich obviňuje [4.1.] a druhý naopak priamo menuje, že tieto práve nevadia [4.5]. Životné prostredie spomínajú iba 4 skupiny autorov. Zameriavajú sa však iba na „nové“ činitele v prírode, ako je DDT, PCB a podobné látky, ktoré majú pôvod v organických uhlíkových zlúčeninách, používaných najmä ako jedy [4.1., 4.2, 4.3, 4.6]. Jedna partia podozrieva elektromagnetický smog [4.6]. Z celkového výsledku je jasné odborné vajatanie, široký záber a nejasný záver.

Jednoznačne ide o pôsobenie podprahových stressorov, ktoré niekde udrú. Nevie sa kedy a prečo. Presne takto sa zdravotne prejavuje aj každé environmentálne riziko. Pôsobí cestou veľmi nízkych dávok, so stochastickým účinkom, je pomalé, obvykle pôsobí až vo vyššom veku. Môže to byť aj pôda a voda, ktorú skúmame my.

2.3. Relatívna úmrtnosť na zhubné nádory prsníka

Sledovali sme 22 ročnú relatívnu úmrtnosť na zhubné nádory prsníka v každej obci Slovenska z pôvodných databáz jednotlivých príčin úmrtí. Indikátor bol pre túto prácu vypočítaný z databázy príčin úmrtí, zbieranej Štatistickým úradom SR v rokoch 1993 až 2008. Dáta za roky 1986 až 1991 sme dostali zo Štátneho zdravotného ústavu v Prahe. Rok 1992 nebol k nájdenu v žiadnej z týchto inštitúcií. Bol preto z práce vynechaný. Do úvahy padli všetky údaje o úmrtí na diagnózu 174 a 175 podľa Medzinárodného katalógu chorôb, 9. revízia a C50 podľa Medzinárodného katalógu chorôb, 10. revízia. Pretože každoročne na túto diagnózu zomiera na Slovensku cca 15 mužov, nerozdelili sme mortalitu na mužov a ženy, ale vzali sme všetko a ako reláciu sme tiež použili všetkých obyvateľov, mužov aj ženy. Prípady úmrtí boli skutočné, počty obyvateľov síce tiež, ale javia sa ako tzv. osoboroky (persons years). Vyšlo pomerne nízke relatívne číslo vzhľadom na mužov vo výpočtoch, ale pre princíp vzťahov a porovnania to nemá význam. Hodnotíme rozdiel, nie absolútny stav. Spolu bolo v analýze 2883 slovenských obcí [5].

2.4. Metodika výpočtu

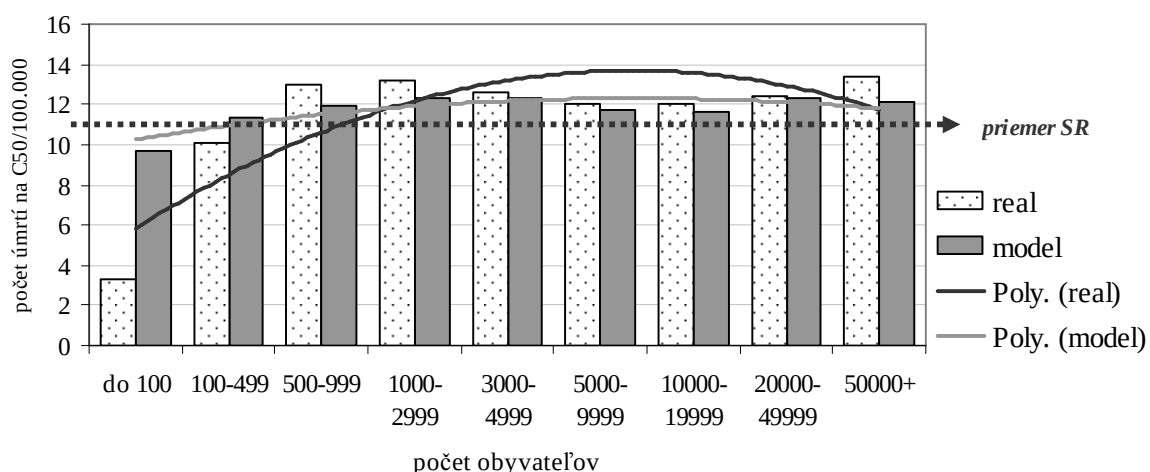
Matematickým aparátom boli neurónové siete z produktu STATISTICA [6]. Metodika tréňovaných neurónových sietí je veľmi zložitá, je z kategórie umelej inteligencie. Krátko

povedané sa hľadajú najvýhodnejšie vzájomné väzby medzi relatívnou mortalitou na rakovinu prsníka (C50) a hodnotou environmentálnych indikátorov. V tomto prípade bol použitý 68-rozmerný model. Každá nasledujúca neurónová sieť je lepšia ako predošlá, pretože sa z predchádzajúcich niečo naučila – „natrénovala“. Výsledkom nemusí byť tá posledná sieť, ktorá je technicky najpokročilejšia, najlepšia môže byť aj niektorá z predošlých. Preto ich musí byť vypočítaných veľmi veľa, aby sa zistilo, kde je ten najlepší moment pre dané riešenie. Kvalita nájdených väzieb sa testuje spätným modelovaním hodnoty relatívnej úmrtnosti na C50 v obciach iba z vybraných indikátorov a vzájomnej funkcie indikátora a mortality. V postupe sa rieši ktoré environmentálne indikátory sú vplyvné a ktoré nie a ako vyzerá funkcia vzťahu u vybraných. Tento súhrn informácií sa napokon prejaví aj v korelačnom koeficiente siete. Pretože v sieti bolo 196044 konečných vstupných dát, ktoré vznikli z niekoľko desiatok miliónov vstupných údajov, je dosahovaný korelačný koeficient zákonite pomerne nízky. Neznamená však, že bezvýznamný. Napokon bolo vypočítaných 200 neurónových sietí, z ktorých systém vybral jednu ako najlepšiu, s pomerne nízkym korelačným koeficientom. Túto sme aj vyhodnotili. Nízky koeficient môže znamenať, že táto sieť nie je dobrá pre niektoré prípady (obce alebo indikátory). Alebo tiež, že environment nie je rozhodujúcim momentom pre prejavenie sa rakoviny prsníka. Zdá sa však, že skôr prvý predpoklad je správny.

3. Výsledky

Bolo vypočítaných 200 neurónových sietí, z ktorých bola jedna vybratá ako najlepšia z nich (s korelačným koeficientom 0,20634). Niektoré iné zdravotné indikátory, ktoré sa podobne počítali pri inej príležitosti, vykazovali korelačný koeficient oveľa vyšší. Znamená to iba toľko, že niektoré zdravotné indikátory sú takmer na 100 percent závislé na kvalite okolitej vody a pôdy. Čo je výsledok takmer nečakaný. Zjavne treba dávať oveľa väčší pozor na to, čo jeme a pijeme.

Aj v našom prípade je však obvykle očakávaných maximálne 20% vplyvu životného prostredia veľmi nízky odhad. Domnievame sa, že aj na rakovinu prsníka vplýva okolitá príroda oveľa viac, ako sme si doposiaľ uvedomovali.



Obrázok 1: Relatívna úmrtnosť na rakovinu prsníka (C50) a veľkosť obce

Na obrázku 1 sú **biele stĺpce** označené ako *real*, ktoré predstavujú reálnu, zo základných štatistických databáz vypočítanú relatívnu úmrtnosť na rakovinu prsníka (C50) v rokoch 1986 až 2008 v slovenských obciach, rozdelených do skupín podľa klasického demografického triedenia obcí podľa počtu ich trvalých obyvateľov, čiže skrátené podľa ich veľkosti.

Šedé stĺpce (*model*) predstavujú modelovanú hodnotu tejto úmrtnosti podľa vybraných dvanástich vplyvných ekologických indikátorov, ktorá vyšla z vybratej neurónovej siete.

Bodkovaná čierna šípka je priemerná celoslovenská hodnota relatívnej úmrtnosti na C50 v rokoch 1986 až 2008 (11,52 úmrtí na 100.000 obyvateľov).

Čierna čiara je polynomické vyrovnanie trendu reálnych hodnôt.

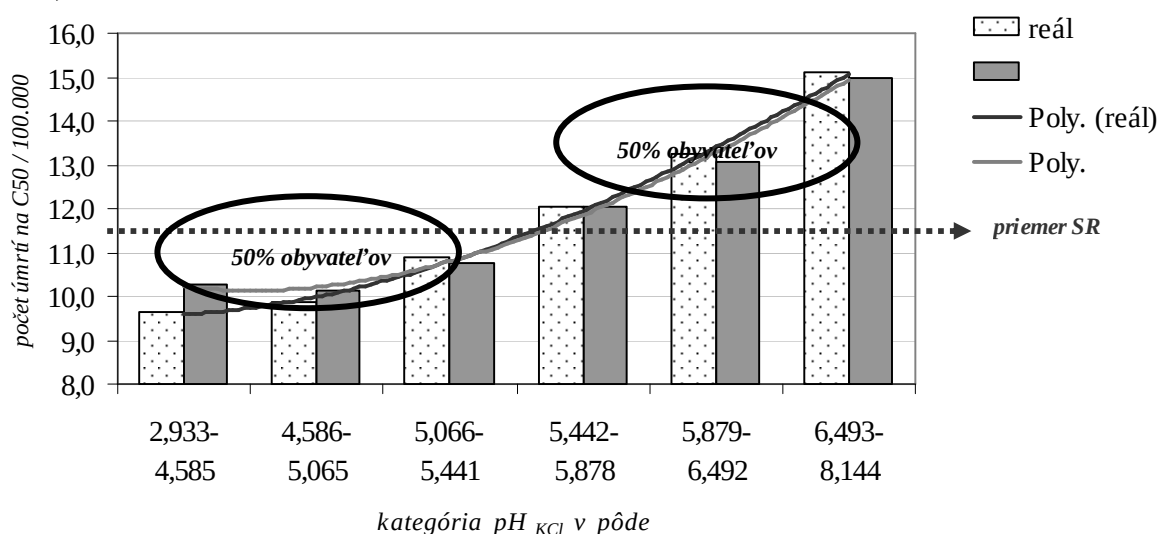
Šedá čiara je polynomickým vyrovnaním modelových hodnôt.

Sledujeme ich preto, aby sa dali porovnať trendy reality a modelu. Vidíme, že nie sú celkom zhodné najmä pre vysoký rozdiel v obciach do 100 obyvateľov. Takýchto obcí bolo v uvedenej dobe na Slovensku 127, čiže sa tento prepočet týka menej ako 1000 obyvateľov Slovenska. V najmenších obciach a ešte aj v ďalšej skupine do 500 obyvateľov dosť často nastala situácia nula – žiadne úmrtie na túto diagnózu v priebehu sledovaných 22 rokov. Túto hodnotu model ani raz nevypočítal, všade príroda pôsobí a nikde nemáme takú oblasť, kde by boli prírodné podmienky natoľko harmonizované, že nevznikne žiadne riziko ochorenia. Pre malé obce je aj 22 rokov príliš krátky čas sledovania. Ak vychádzame z celoslovenskej relatívnej hodnoty 11,52 prípadov na 100.000 obyvateľov, v obci, ktorá má 100 obyvateľov, sa očakáva jedno úmrtie raz za 86 rokov. Tento časový horizont nie sme schopní overiť. Druhá strana – veľké mestá – sú závažnejšou informáciou, ide o oveľa viac ľudí. Možno tu pôsobí viac stresujúcich faktorov ako na vidieku a riziko vzniku rakoviny z iných ako prírodných dôvodov je vyššie aj z tohto dôvodu. Alebo je tu objektívne vyššie zastúpenie odborných lekárov a kvalitnejšia diagnostika, ako aj presnejšie vypisovanie úmrtných listov. Určite tu je takmer nemožné stanoviť konkrétnu geografickú oblasť dopestovania potravín a už vonkoncom sa nedá predpokladať, že toto obyvateľstvo je prevažne zásobované potravinami lokálnej výroby. Tento demografický úvod bol nevyhnutný pre porozumenie výsledkom a voľnosti kvality siete.

3.1. Popis vplyvných ekologických indikátorov

Ako najvplyvnejšie sa ukazuje byť pH_{KCl} v pôde. Tento indikátor je určitým vyjadrením pôdnej kyslosti a jeho hodnota dobre indikuje kationovú výmennú kapacitu a kationové zloženie sorpčného komplexu pôd.

Na obrázku 2 je opäť porovnanie reality a modelu podľa výšky pH_{KCl} v pôde. Kategórie pH_{KCl} boli stanovené podľa počtu obcí v tomto rozsahu. V každej kategórii je 16,66% obcí (presne 481), zotriedené podľa hladiny pH . V tomto prípade sa model aj realita takmer zhodujú až na mierne rozdielny prvý stĺpec pri najkyslejších pôdach. Polovica obcí, ktorá existuje na kyslých pôdach, má aj najnižšiu relatívnu úmrtnosť na rakovinu prsníka. Obce s pôdami s iba miernou kyslosťou, neutrálnymi až alkalickými majú oveľa vyššiu túto mortalitu. V rámci 2883 obcí bola najnižšia hodnota pH_{KCl} v pôde rovná 2,96 (Smolník v okrese Gelnica) a najvyššia hodnota pH_{KCl} v pôde rovná 8,14 (Nová Dedinka v okrese Senec).



Obrázok 2: pH_{KCl} v pôde a relatívna úmrtnosť na rakovinu prsníka

Ako vplyvné prvky vyšli nasledovné environmentálne indikátory, ktoré sú uvedené v tabuľke zostupne podľa výšky ich vplyvu.

Tabuľka 1: Vplyvné environmentálne prvky na vznik rakoviny prsníka

poradie podľa vplyvu	nosné prostredie	prvok - substancia	miera vplyvu	popis funkcie, limity
1	pôda	pH _{KCl}	1,000907	s rastom pH _{KCl} stúpa mortalita na C50, optimálna hladina je pH _{KCl} < 4,47, kritická hladina 7,00
2	voda	Ba	1,000385	čo najmenej, max. Ba ≤ 0,04, kritická hladina 0,39
3	voda	SiO ₂	1,000314	menej, do 27,9, kritická hladina 62
4	voda	Zn	1,000279	menej, do 0,24, kritická hladina 2,21
5	voda	Fe	1,000272	čo najmenej, kritická hladina 1,9
6	pôda	Ce	1,000271	náznak paraboly, optimum 44,9 – 51,8
7	pôda	Mg	1,000230	čo najmenej, kritická hladina 2,83
8	pôda	karbonáty	1,00023	čo najmenej, kritická hladina 28
9	pôda	Na	1,000141	náznak paraboly, optimum 0,71 – 1,02
10	pôda	Ca	1,00013	čo najmenej, kritická hladina 7,00
11	pôda	Co	1,00010	náznak paraboly, optimum 7,25 – 11,55
12	pôda	Sr	1,00003	menej, do 84,9, kritická hladina 281

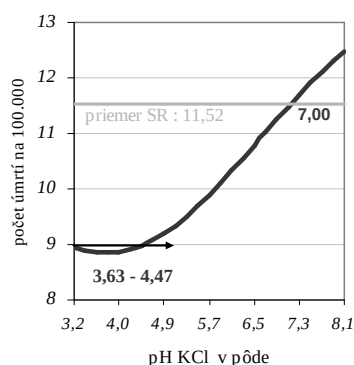
Ešte ďalších 20 prvkov bolo označených ako síce vplývajúce, ale iba veľmi málo (koeficient vplyvu pod 1,0000) a zvyšných 35 prvkov bolo vyradených ako celkom nevplývajúce.

Výsledky predvedieme formou zhody späťne modelovaných hodnôt relatívnej úmrtnosti na C50 so skutočnými v rámci kategórie obsahu prvku v nosnom prostredí. Úvodným obrázkom je skúška správnosti výpočtu - zhoda modelu (*výpočet relatívnej úmrtnosti na C50 iba podľa funkcií vplyvu vybraných environmentálnych prvkov*) s realitou (*relatívna úmrtnosť na C50 v rokoch 1986-2008*) podľa veľkosti obce.

Najhoršie sa modelovali celkom malé obce, ktoré majú výrazne deformované vekové proporcie a následkom toho aj všetky demografické a zdravotné udalosti dostávajú extrémne hodnoty v relatívnom pojatí.

Obce, ktoré majú 3000 až 50000 obyvateľov, sú si zrejme podobné v životnom štýle aj vekovom zložení. V týchto obciach sa realita od modelu líši iba minimálne. Žije v nich cca 40% obyvateľstva SR. V obciach do cca 500 obyvateľov je model nadhodnotený, vo veľkostiach podhodnotený. Zrejme pre tieto problémy bol korelačný koeficient tohto modelu pomerne slabý a sieť nebola vyhodnotená ako príliš významná. Napriek tomu je až zarážajúce, s akou presnosťou sa dá urobiť model mortality na rakovinu prsníka v obci iba zo vstupov, ktoré sú predstavované čisto ekologickými indikátormi z podzemnej vody a pôdy.

Polovica obyvateľstva žije na výrazne kyslých pôdach a má zároveň aj nízku mortalitu na rakovinu prsníka. Všetky tieto obce sú v realite aj v modeli pod celoslovenským priemerom v dlhodobej mortalite na C50. Podobný výsledok vyšiel aj pri iných typoch zhubných nádorov. Druhá polovica žije na pôdach menej kyslých, neutrálnych až alkalických. Rozdiel v neprospech alkalických pôd je až o 50% vyššia sledovaná mortalita. Tento výsledok je treba prinajmenšom overiť samostatnou štúdiou. Toto je iba strohý popis skutočnosti.

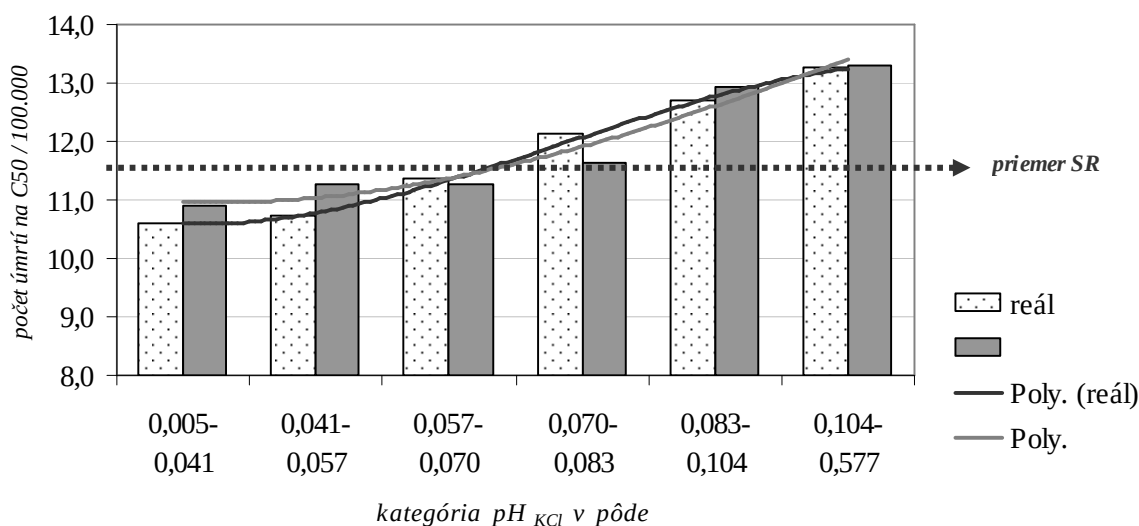


Zdá sa, že je výhodné, ak je pH_{KCl} v pôde nižšie. Podľa našich výpočtov je optimálna hladina do 4,47, čiže kyslá pôdna reakcia. Počnúc touto hodnotou už relatívna úmrtnosť na rakovinu prsníka so zvyšovaním tohto environmentálneho indikátora strmo stúpa. Pri hodnote $\text{pH}_{\text{KCl}} = 7,00$ už pretína hodnotu priemeru SR. Podľa publikovaných prác má mať kvalitná pôda neutrálnu, alebo mierne zásaditú reakciu [7].

Obrázok 3: pH KCl v pôde a rakovina prsníka

Zdá sa, že táto premisa neplatí pre rakovinu prsníka, skôr naopak. Podobný výsledok sme zaznamenali aj pri relatívnej úmrtnosti na všetky druhy zhubných nádorov bez rozlíšenia druhu. Tu sme našli optimálnu hladinu pre najnižšiu úmrtnosť v rozmedzí 5,3 až 5,5, čiže taktiež kyslá pôdna reakcia. Na základe tohto usudzujeme, že nejde o náhodný výsledok, ale že mierne kyslá pôda je skutočne mierne protektívna. Resp. naopak, alkalická pôda je výrazne riziková.

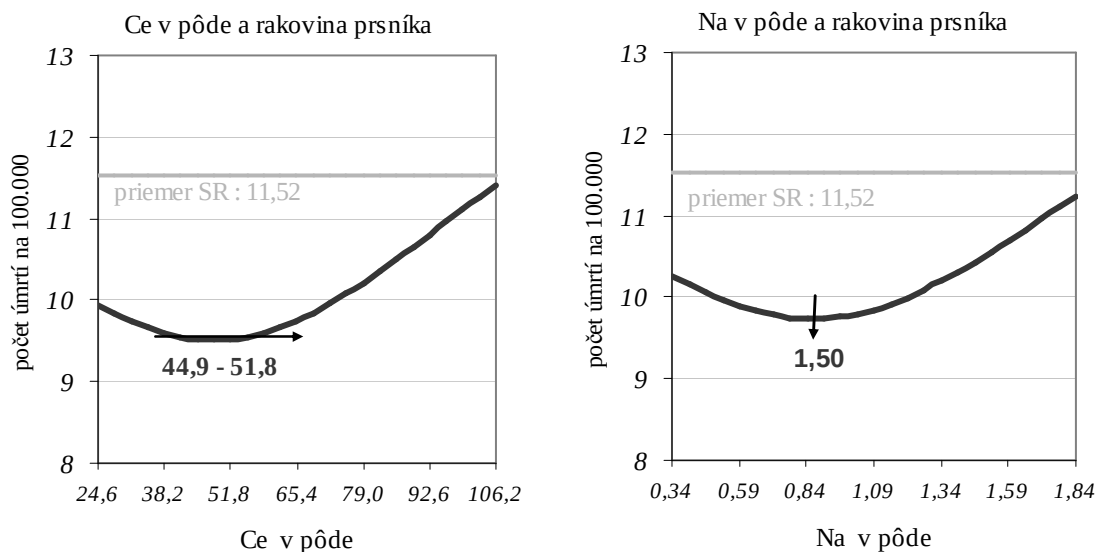
Druhým najvplyvnejším indikátorom je bárium vo vode. **Bárium** (*barium*) patrí medzi prvky alkalických zemín, je mäkký, reaktívny kov. Búrlivo reaguje s kyslíkom aj vodou, a preto sa s ním v prírode stretávame len v podobe zlúčenín. Ťaží sa aj na Slovensku ako nerast barytu. Je zrejme vysoko toxické, jeho rôzne zlúčeniny sa používajú ako súčasť pyrotechniky, otravy na potkany, pri röntgenovom snímkaní tráviaceho traktu vypije pacient suspenziu síranu bárnateho vo vode a po niekoľkých desiatkach minút možno získať kvalitnú snímku žalúdka a čriev pacienta. Tiež je súčasťou omietok a bielej farby. V morskej vode je v priemernej koncentrácii 0,03 mg/l. [8]. Jeho negatívne vlastnosti potvrdzujeme, javí sa nám ako nebezpečné aj v najnižších koncentráciách.



Obrázok 4: Bárium vo vode a relatívna úmrtnosť na rakovinu prsníka

Z ostatných prvkov vyberáme zaujímavosti:

Cér aj **nátrium** v pôde sa zdajú byť protektívne, určitá hladina je zrejme žiaduca.

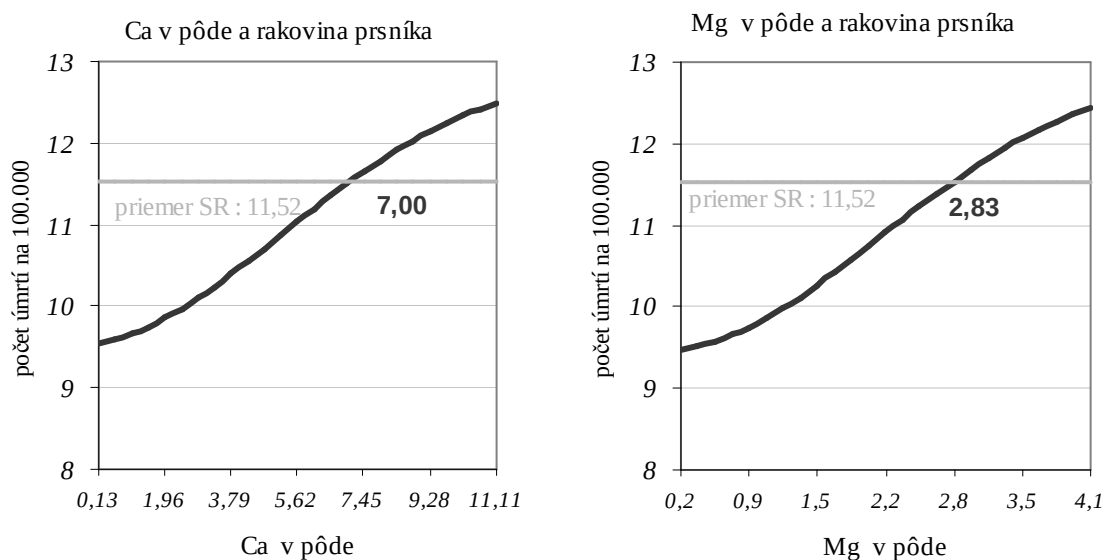


Obrázok 5: vybrané prvky, ktoré sú účinne, ak majú určitú hladinu

Medzi prvky s požadovanou hladinou ešte patrí **kobalt v pôde**. Nátrium potvrdzuje teóriu protektivity slaných pôd. Kobalt je zaujímavý, možno je to prejav vitamínu B12, ktorý potrebuje práve kobalt k svojmu pôsobeniu.

Naopak, **zinok, kalcium, železo a magnézium** sa javia ako nebezpečné a s ich rastom takmer lineárne stúpa aj mortalita na C50.

Tieto látky sa javia ako nebezpečné látky v akomkoľvek množstve, čo napokon potvrdzuje teóriu vyššej kyslosti pôd, ktorá sa ukázala byť ako najzávažnejší environmentálny moment.

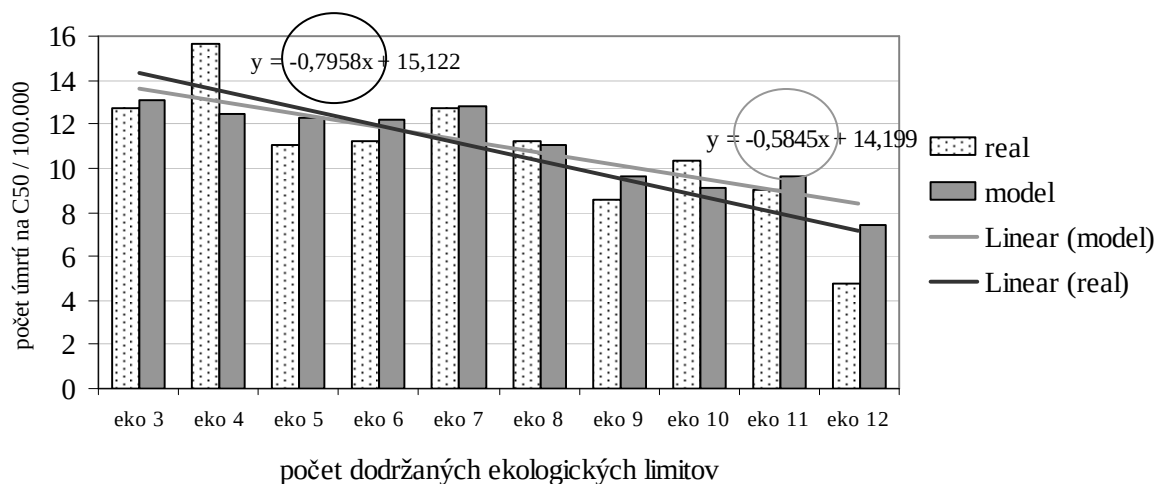


Obrázok 6: vybrané prvky, ktoré sú nebezpečné v akomkoľvek množstve

Porovnali sme tiež rozdiely v relatívnej mortalite na C50 v obciach, kde boli a neboli dodržané nami navrhované environmentálne limity. Spolu ide o 12 prvkov s odporučenými optimálnymi hodnotami. Všetkých 12 limitov bolo dodržaných v troch obciach, kde bola modelovaná mortalita 7,41 úmrtí na 100.000 obyvateľov. Reálna bola 4,81.

Na obrázku č.7 je táto informácia v stĺpci označenom ako *eko 12*. Nie je ani jedna obec, kde by neboli dodržané aspoň 2 nami navrhované limity. S rastúcim počtom dodržaných limitov klesá relatívna mortalita na C50 ako v realite, tak v modeli.

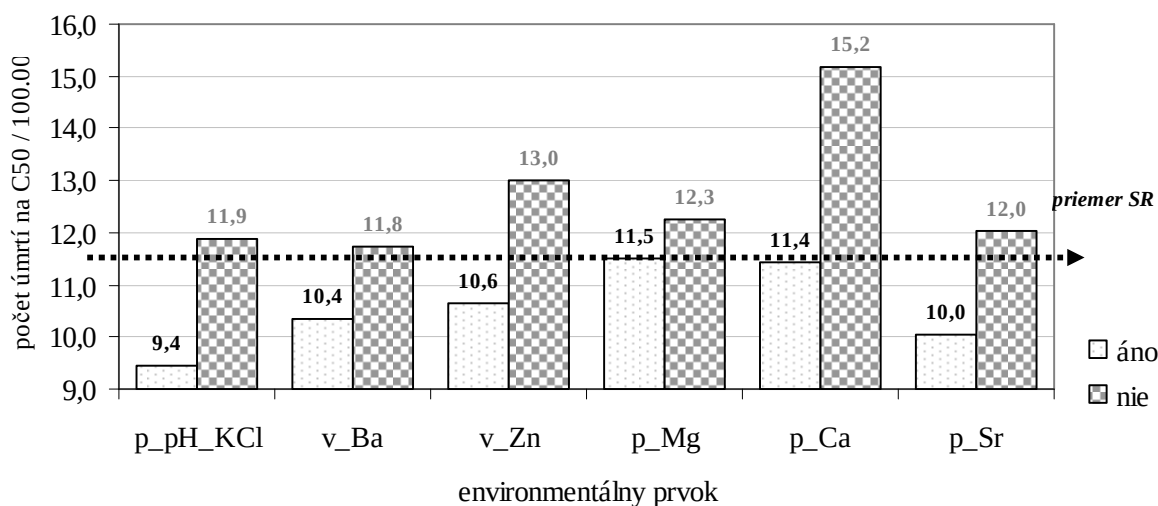
V realite to vyzerá tak, že s každým dodržaným našim ekologickým limitom klesá počet úmrtí na 100.000 obyvateľov o 0,79 prípadov, v modeli je to o 0,58 prípadov, čiže reálny stav je ešte viac viazaný na životné prostredie ako modelový.



Obrázok 7:

Počet dodržaných ekologických limitov a relatívna úmrtnosť na rakovinu prsníka

Pri testoch sa ukázali byť preukazné rozdiely v mortalite medzi obcami, kde boli a neboli dodržané navrhované limity, v šiestich environmentálnych indikátoroch. Tieto indikátory sú prehľadne na nasledujúcom obrázku číslo 8.



Obrázok 8: Preukazne rozdielna úmrtnosť na rakovinu prsníka pri dodržaní limitov

Ostatné prvky (SiO_2 vo vode, Fe vo vode, Ce v pôde, karbonáty v pôde a Co v pôde) nezaznamenali štatisticky preukazný rozdiel.

Avšak tých šesť environmentálnych indikátorov, ktoré sú na obrázku č.8, jednoznačne vplýva na úmrtnosť na rakovinu prsníka. Navrhované limity sa zdajú byť dôležité. Netrúfame si však povedať, že ostatných šesť je bezvýznamných, keďže ich sieť tiež vybrala ako vplývajúce.

4. Záver

Pracovali sme iba s databázami, ktoré garantuje štát. Tieto dáta sú dlhodobo zbierané jednotnou, medzinárodne akceptovanou metodikou, prostredníctvom špecializovaných odborných štátnych inštitúcií, ktorými sú štatistický úrad Slovenskej republiky a Štátny geologický ústav. K výpočtom a hodnoteniu sme používali iba profesionálne databázové a štatistické programové balíky. Do tohto seriózneho systému sme vniesli iba nápad. Zisťovali sme, či spolu súvisia environmentálne indikátory a mortalita na rakovinu prsníka, nech už je sprostredkujúci medzičlen akýkoľvek. Preto sme popísaným spôsobom vypočítali priemernú hodnotu 67 environmentálnych indikátorov a relatívnu hodnotu úmrtnosti na rakovinu prsníka pre každú z 2883 obcí Slovenska. Týchto 68 x 2883 údajov sme vložili do trénovanej neurónovej siete, z ktorej po určitej dobe vyšla najvýhodnejšia. Táto sieť určila, že 55 indikátorov neovplyvňuje v dostatočnej miere rakovinu prsníka. Naopak, 12 indikátorov ju preukazne ovplyvňuje. Sú to tieto (v zostupnej miere od prvého po posledný): pH_{KCl} v pôde, Ba vo vode, SiO_2 vo vode, Zn vo vode, Fe vo vode, ostatné sú v pôde: Ce, Mg, karbonáty, Na, Ca, Co a Sr. Pre všetky tieto indikátory boli skonštruované funkcie, v ktorých sa najlepšie zachytáva ich vzťah k mortalite na rakovinu prsníka. Pre cer, natrium a kobalt vyšla parabola – je žiaduca optimálna hladina. Pre ostatné vyšla takmer lineárna funkcia, ktorá predpokladá najnižšiu hladinu indikátora pri najnižšej mortalite („čím menej, tým lepšie“).

Pri overovaní modelu sme prišli k záveru, že iba prostredníctvom týchto 12 funkcií sa dá pomerne presne modelovať reálna mortalita na rakovinu prsníka v konkrétnej obci.

Výnimku tvoria malé obce do 100 obyvateľov, kde je všetko relativizované vždy prehnané následkom chyby malých čísel. Nie najpresnejšie vyšli aj mestá nad 50.000 obyvateľov. Je takmer isté, že vo veľkých mestách je už takmer nemožné vystopovať zdroje potravín a dá sa reálne predpokladať, že iba veľmi málo súvisia s lokálnou produkciou. Preto je tu veľmi ťažké dokázať vplyv lokálneho znečistenia pôdy, snád' iba vody. Obe tieto demograficky podmienené rozkolísania iba potvrdzujú správnosť modelovania.

Jedným z výsledkov týchto funkcií je aj možnosť odhadu limitných hodnôt. Teda takých, pri ktorých hodnota mortality nestúpa, ale naopak klesá, alebo aspoň stagnuje. Za kritické hladiny sme považovali tie, kde hodnota environmentálneho indikátora prekračuje dlhodobý celoslovenský priemer v mortalite na C50. Potom sme testovali rozdiel v realite aj v modeli medzi obcami, kde boli dodržané tieto navrhované limity a kde neboli dodržané.

Štatisticky preukazný rozdiel medzi reálnou hodnotou mortality a modelovanou sa našiel u šiestich environmentálnych indikátorov. Vo všetkých prípadoch bola realita aj model preukazne nižšia v obciach, kde bol tento limit dodržaný. Tiež bol preukazný rozdiel medzi obcami, kde bol dodržaný vyšší počet limitov. Každý dodržaný limit znižoval v obci mortalitu na rakovinu prsníka o 0,79 prípadov na 100.000 obyvateľov.

Šesť vysoko preukazne vplývajúcich environmentálnych indikátorov je: pH_{KCl} v pôde, bárium a zinok vo vode, magnézium, kalcium a stroncium v pôde. Pri dodržaní limitu v obci bola dlhodobá mortalita na rakovinu prsníka preukazne nižšia.

Je nepresné hovoriť o nejakom dodržaní limitu či nedodržaní. Ide o prírodný jav, ktorý človek asi rýchlo neovplyvní a preto nemôže jednoduchým spôsobom niečo dodržať či nie, ale nenašli sme lepšie slovné vyjadrenie. V mnohých prípadoch ide o antropogénne znečistenie, či aj zavlečenie cudzích substancií v dobrom úmysle (hnojenie).

Uvedomujeme si, že všetky vstupné dáta boli určitým spôsobom priemerované alebo inak ošetrované a že sú teda pomerne nepresné. Tiež metodika výpočtu je pomerne voľná. Pri výpočte ďalších tisícok neurónových sietí by možno vyšlo aj niečo iné. Čiže aj vstup aj výstup je nepresný. Avšak je postavený na obrovskom dátovom základe a aj tieto výsledky sú veľmi presvedčivé.

Preto ako úplný záver hovoríme, že by to mohla byť celkom ľahko aj pravda a prinajmenšom je potrebné výsledné skutočnosti podrobiť ďalším podrobným skúmaniam. Možno to je stopa.

5. Literatúra

- [1]RAPANT, S., LETKOVIČOVÁ, M., CVEČKOVÁ, V., GALBAVÝ, J.,: Environmentálne a zdravotné indikátory Slovenska, manuscript, ŠGÚDŠ 2010, 276 str.
- [2]METODICKÝ POKYN MŽP SR č. 632/98
- [3]VRANA, K., RAPANT, S., BODIŠ, D., MARSINA, K., LEXA, J., PRAMUKA, S., MAŇKOVSKÁ, B., ČURLÍK, J., ŠEFČÍK, P., VOJTAŠ, J., DANIEL, J. & LUČIVJANSKÝ, L., 1997: Geochemical Atlas of Slovak Republic at a scale 1 : 1 000 000. In: Jurnal of Geochem. Exploration, 60, s. 7 – 37.
- [4]*súbor 10 vybratých skupín autorov v literárnej rešerši príčin rakoviny prsníka:*
 1. Overview: Breast Cancer
What Causes Breast Cancer?
http://www.cancer.org/docroot/cric/content/cric_2_2_2x_what_causes_breast_cancer_5.asp
American cancer Society
Last Medical Review: 09/29/2009
 2. http://www.ehealthmd.com/library/breastcancer/brc_causes.html
ehealthMD.com
 3. <http://www.cancerhelp.org.uk/type/breast-cancer/about/risks/possible-breast-cancer-risks#pesticides>
Cancer Research UK is the leading funder of cancer research in the UK
 4. Do We Know What Causes Breast Cancer?
http://www.acsevents.org/docroot/CRI/content/CRI_2_4_2X_Do_we_know_what_causes_breast_cancer_5.asp?rna v=cric
 5. <http://www.nlm.nih.gov/medlineplus/ency/article/000913.htm>
Causes
www.cancer.gov/bcrisktool
 6. http://www.thenutritionreporter.com/breast_cancer.html
Tracing the Elusive Causes of Breast Cancer And Leveraging Dietary Strategies to Prevent It
 7. What Causes Breast Cancer?
WebMD Health News
<http://www.webmd.com/breast-cancer/guide/what-causes-breast-cancer>
 8. <http://www.mayoclinic.com/health/breast-cancer/DS00328/DSECTION=risk-factors>
Causes
By Mayo Clinic staff. Living with cancer newsletter
Subscribe to our Living with cancer newsletter to stay up to date on cancer topics.
 9. Breast Cancer
(Mammary Cancer, Cancer of the Breast)
http://chealth.canoe.ca/channel_condition_info_details.asp?disease_id=23&channel_id=12&relation_id=1619
 10. Breast Cancer - What Increases Your Risk
By Kathe Gallagher, MSW
MORE ON BREAST CANCER
Medical Review: Joy Melnikow, MD, MPH - Family Medicine Douglas A. Stewart, MD - Medical Oncology
Last Updated: 08/31/2007
<http://health.yahoo.com/breastcancer-causes/breast-cancer-what-increases-your-risk/healthwise--tv3625.html>
- [5]Vekové zloženie obcí 1993 až 2008, Individuálne príčiny úmrtí 1993 až 2008, Štatistický úrad SR, Individuálne príčiny úmrtí 1986 až 1991, Státní zdravotní ústav, Praha
- [6]Statsoft, STATISTICA 7, Neural Network
- [7]J.ČÉRY, J. ÁRVAY, R.STANOVIČ: Monitoring hygieny poľnohospodárskej pôdy v regióne Brekov. In: 8. vedecká konferencia doktorandov a mladých vedeckých pracovníkov, 18.-19.4.2007, Nitra, FPV UKF Nitra.
- [8]<http://sk.wikipedia.org/wiki/B%C3%A1rium>

Adresa autorov:

Mária Letkovičová, Ing., CSc.
Jaskyňová 3
949 01 Nitra
letkovicova@environment.sk

Hana Zach, Ing., PhD
Čajkovského 11
949 11 Nitra
zach_hana@yahoo.co.uk

Priestorová štúdia rozmiestnenia rýchlej zdravotnej pomoci Space study of allocation of health rescue services

Mária Letkovičová a Hana Zach

Abstract

We have analyzed all the municipalities of the Trnava Autonomous Region and all the neighboring municipalities of other regions. For each village we have calculated time series of relative mortality rates for selected diagnoses for which is commonly doctor called (years 1994-2008). As the calculation we used the fuzzy c-cluster analysis of the status and trends of relative mortality for selected diagnoses. We have found that coherent areas with high status and increasing trend can be found where there are natural or artificial terrain obstacles such as mountains, waterways, highways etc. on the road in the direction of the emergency ambulance services. If only one village is marked like this, there is located one of the institution of social welfare such as retirements home. We recommend this procedure for the spatial analyses of the efficiency of emergency services and quality of social welfare institutions.

Key words: mortality, fuzzy c-cluster analysis, spatial analysis, GIS, health rescue services

Kľúčové slová: úmrtnosť, fuzzy c-zhluková analýza, priestorová analýza, GIS, zdravotné záchranné služby

1. Úvod

Sociálna a zdravotná politika je spojením potreby pomoci s možnosťou dávať. Toto spojenie býva chápané v celkom inej rovine ako matematickej. Nielen humanitnej, o čom nás presvedčuje celý rad škandálov, ktorý v poslednej dobe sprevádza spoločenskú stránku vzájomnej kompenzácie. Omyl. Všetko má svoje celkom dobré optimalizačné metódy a doslova "všetko vieme". My štatistici. Možno by bolo dobré viac analyzovať a menej súťažiť. Peniaze nie sú všetko. Čo napokon vie každý, kto na sanitku čakal príliš dlho.

Predkladaný príspevok je výňatkem zo štúdie, v ktorej sa urobila súhrnná priestorová analýza viacerých skutočností v jednej lokalite [1]. Dávali sa do súvisu geografické podmienky, existujúce dopravné možnosti a špecifická úmrtnosť. Výsledkom môže byť napríklad aj posúdenie správnosti rozmiestnenia záchranej zdravotnej služby a jej efektívnosť.

2. Cieľ práce

Cieľom štúdie, z ktorej je urobený tento príspevok, bolo vyhodnotiť efektívnosť rýchlej záchranej zdravotníckej služby v kraji Trnava. Urobili sme to pomocou priestorovej analýzy relatívnej úmrtnosti na tzv. „rýchle“ diagnózy. Takto sme pomenovali také, ku ktorým obvykle býva volaná niektorá z foriem rýchlej zdravotníckej pomoci. Predpokladalo sa, že je možné, že tam, kde je táto úmrtnosť nižšia, pracuje záchranná služba efektívnejšie. Jedno z možných vysvetlení tiež je, že sa k problému dostane včas. Pomocou štatistických metód sa hľadali miesta, ktoré pravdepodobne vykazujú ťažšiu dopravnú dostupnosť. Jedným z cieľov bol aj pokus o vyhodnotenie účinnosti zmeneného systému zdravotnej záchranej služby v rokoch, ktorý pracoval v rokoch 2006 a neskôr. Porovnávali sa roky 2004+2005 s rokmi 2007+2008. Po novom konkurze na inú správu záchranných systémov sme dostali možnosť urobiť čo nevidieť tretie porovnanie.

Celkom konkrétnym cieľom hodnotenia bolo zistiť:

- Či sa v niektorých oblastiach Kraja vytvárajú oblasti s odlišným stavom alebo odlišným trendom podielu úmrtí na vybrané diagnózy, ku ktorým obvykle býva volaná zdravotná pomoc. Identifikovaná odlišná oblasť je základom pre ďalší postup.

- Či tieto odlišnosti presahujú rámec Kraja – sú všeobecne platné pre väčšie územie, čiže asi majú globálne príčiny, ťažko odstrániteľné zásahom jedného Kraja, alebo sa týkajú iba Kraja a potom je riešenie v jeho kompetencii a dá sa povedať, že aj v jeho povinnostiach.

3. Dáta a metódy

Všetky podkladové dáta boli zakúpené v Informačnom servise Štatistického úradu SR buď vo forme individuálnych záznamov o úmrtí, alebo podrobnej vekovej štruktúry obcí [2]. Boli použité ako klasické štatistické metódy, tak aj moderné priestorové analýzy. Základné softvérové vybavenie pozostávalo najmä z geografického informačného systému ARC VIEW, kde sa brali do úvahy najmä vrstvy Cesty a Vodstvo (okrem hraníc obcí, okresov a krajov), niekoľkých databázových a štatistických profesionálnych balíkov (EpiInfo, Syn-Tax) ako aj z vlastných účelovo zhotovených programov vo FoxPro.

Analýzovala sa úmrtnosť na vybrané diagnózy, ktoré pozostávali z infarktov a porážiek v rámci prirodzených príčin a všetkých druhov násilných úmrtí, v prepočte na 100.000 obyvateľov (skrátene nazvaná **výber**). Sledovali sa časové rady samostatne pre každú obec, pre každý okres a pre každú z troch častí VÚC Trnava, ako aj pre celé Slovensko v relevantnej dobe. Hlavnou výpočtovou technikou bola fuzzy c-zhluková analýza, v ktorej sa modeloval slovný pojem: dať do jedného zhluku podobný stav a trend v relatívnej úmrtnosti na vybrané diagnózy [3]. Výsledkom bol vždy stanovený najvýhodnejší počet zhlukov, na ktorý sa súbor dát rozpadá, stredy zhlukov, ktoré po spojení dávajú graf trendu a funkcie príslušnosti každej obce ku každému zo zhlukov. Ďalej sme špeciálne zohľadňovali vekovú proporcionalitu v obci s dôrazom na podiel starších obyvateľov a modelovali očakávanú úmrtnosť vzhľadom na tento podiel [4].

Relatívna úmrtnosť na vybrané diagnózy bola sledovaná v rámci 5-ročných prekrývajúcich sa období. Takto zvolený čas už pomerne slušne eliminuje prípadné extrémne roky, alebo tiež vplyv malých čísel v malých obciach. Pre celkom malé obce je aj 5 rokov krátky čas, ale vo VÚC Trnava a jeho okolí celkom malé obce vôbec nie sú.

Priestorové porovnanie stavu a trendu zabezpečujeme sledovaním obcí v susediacich okresoch. Buď všetkých obcí, pokiaľ ide o malý okres, alebo výberu okresu, ktorý pozostáva z obcí, ktoré susedia s okresmi VÚC Trnava.

Sledujeme lokálne súvislosti a vplyvy, ktoré obvykle nekončia hospodárskymi hranicami kraja. Zisťujeme, či je daný problém v celej geograficky blízkej lokalite, čiže je daný miestnymi životnými podmienkami, alebo či je daný pravidlami, ktoré platia iba pre konkrétne hospodárske celky.

Ak je v celej Podunajskej nížine vysoká vybraná úmrtnosť, je to globálny problém, ktorý jeden okres nevyrieši. Ak je táto úmrtnosť vysoká iba v skupine niekoľkých susediacich obcí, je to podnet na zamyslenie pre miestnych zodpovedných.

Preto „nestriháme“ analýzu presne hranicami okresov a kraja, ale mierne ju cez ne presahujeme. Analyzujeme, ako to vyzerá u susedov. Pravdepodobne to bude tak vyzerat' aj „doma“. Spolu je v analýze zahrnutých 509 obcí, ktoré patria do VÚC Trnava a priľahlých susediacich okresov. Týchto 509 obcí predstavuje cca 22% obyvateľstva Slovenska. Obce VÚC Trnava a susediace okresy sú rozdelené približne na polovicu: 50% obyvateľstva v analýze je z VÚC Trnava a druhých 50% predstavujú obyvatelia okolitých okresov.

Trnavský samosprávny kraj má tvar, ktorý priamo ponúka, aby bol rozdelený na tri časti, ktoré sú aj národnostne aj geograficky odlišné.

Analýzujeme každú časť samostatne v rámci jej susediacich okresov.

Základné rozdelenie analýz je na tieto celky:

- **Sever – okresy na Záhorí:** Senica a Skalica (kopcovitý až hornatý terén, často s charakterom Kopaníc, pieskové pôdy, obyvateľstvo s ladením blízkym Morave, hlavným problémom je horský terén, najmä Karpaty, a diaľnica)

• **Stred – okresy v centrálnom Trnavsku:** Trnava, Piešťany a Hlohovec (rovina, Trnavská tabuľa, slovenské obyvateľstvo je v absolútnej väčšine, hlavným problémom sú Karpaty a diaľnica)

• **Juh – okresy na juhu kraja:** Dunajská Streda a Galanta (výrazná rovina, ktorá hraničí s Maďarskom, obyvateľstvo s ladením blízky Maďarsku, hlavným problémom je Dunaj a množstvo zavlažovacích kanálov)

Vybrané diagnózy, ktoré sú zahrnuté do výberu sú nasledovné:

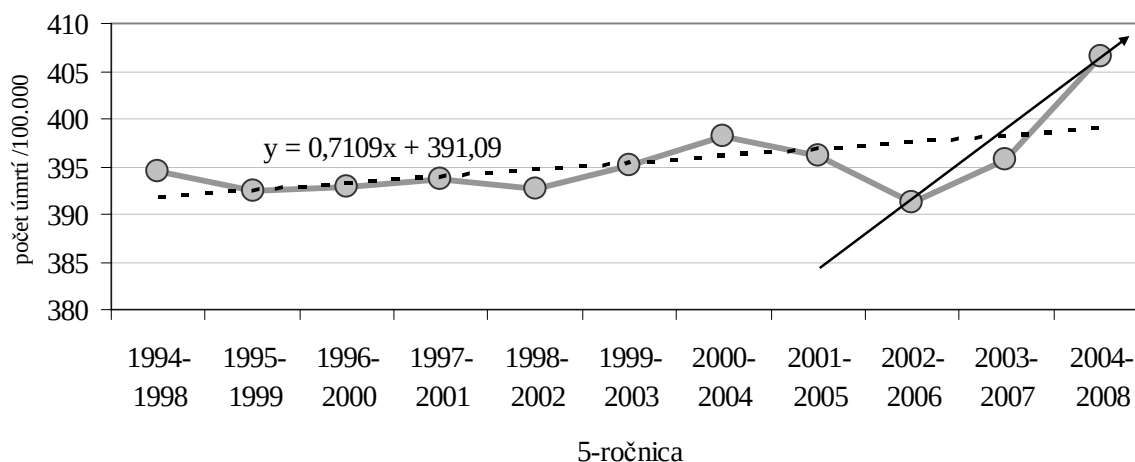
- Ischemické choroby srdca
 - I21 Akútny infarkt myokardu
 - I22 Ďalší infarkt myokardu
 - I23 Doktoré terajšie komplikácie po akútnom infarkte myokardu
 - I24 Iné akútne formy ischemickej choroby srdca
 - I25 Chronická ischemická choroba srdca
- Cievne choroby mozgu
 - I63 Mozgový infarkt
 - I64 Porážka-apoplexia, nešpecifikovaná ako krvácanie alebo infarkt
- Poranenia, otravy a doktoré iné následky vonkajších príčin S00 – T98

Kódové označenie a presné názvy diagnóz pochádzajú z Medzinárodnej štatistickej klasifikácie chorôb a príbuzných zdravotných problémov MKCH-10 [5].

4. Výsledky a hodnotenie

4.1. Celoslovenské výsledky

Podrobné hodnotenie územia sme vykonali pomocou fuzzy c-zhlukovej analýzy. Avšak pre celé Slovensko ako štandard sme zvolili priamy údaj. Na obrázku 1 je vidieť, že v ostatných rokoch prichádza na celom Slovensku k prudkému nárastu úmrtnosti na vybrané diagnózy. Po dlhšom období stagnácie zhruba do roku 2002, kedy indikátor varíroval na hodnote 392-394 prípadov na 100.000 obyvateľov, prichádza k rozkolísaniu stavu a k jeho rýchlemu rastu k hodnote cca 406 prípadov. Toto relatívne malé zvýšenie znamená absolútny nárast o cca 700 prípadov ročne, čo je napríklad ročný počet úmrtí na rakovinu prsníka, alebo štvornásobok ročného počtu na rakovinu kľčka maternice, ktoré sú oboje veľmi medializované ako spoločensky závažný jav. V rámci regresnej rovnici trendu sa javí ako nárast 0,7 prípadu na 100.000 obyvateľov ročne.



Obrázok 1: Relatívna úmrtnosť na vybrané príčiny na Slovensku v rokoch 1994 až 2008

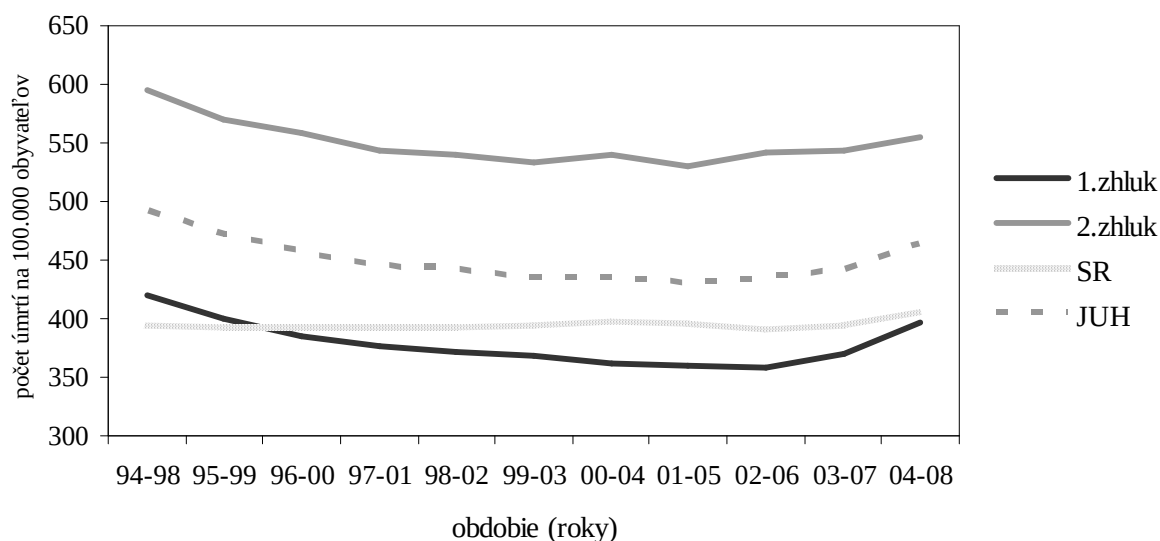
4.2. Výsledky a hodnotenie indikátora VÝBER diagnóz v južnej oblasti Trnavského samosprávneho kraja ako ukážka podrobnej analýzy jednej oblasti

Označujeme ho skratkou ako JUH VÚC TT, alebo iba JUH. Analyzuje trnavské okresy Dunajská Streda (67 obcí) a Galanta (36 obcí). Ďalej zahrnuje na východe susediace okresy z VÚC Nitra: Komárno (40 obcí), Nitra (62 obcí) a Šaľa (11 obcí). Spolu je v tejto analýze zahrnutých 217 obcí. V tabuľke 1 sú všetky výsledky zhrnuté v stĺpci prvý a druhý zhluk, ako aj porovnanie s relevantným stavom na Slovensku (*quasi štandard*) a stavom sledovaného územného celku bez rozdelenia na zhluky. V poslednom stĺpci je percentuálny rozdiel južnej oblasti trnavského kraja (JUH) voči celoslovenskému štandardu. V každom časovom období je na tom Slovensko lepšie ako táto oblasť.

Tabuľka 1: Hodnoty indikátora „výber“ v oblasti JUH VÚC TT

Stred VÚC TT	počet úmrtí na 100.000 obyvateľov				SR voči JUH VÚC TT je nižšia (%)
	centroidy fuzzy c-zhlukovej analýzy		porovnateľné územia		
5-ročnica	1.zhluk	2.zhluk	SR	JUH	
1994-1998	419,9	595,4	394,5	492,4	-24,79
1995-1999	400,4	570,8	392,5	471,3	-20,07
1996-2000	385,1	558,7	392,8	457,0	-16,33
1997-2001	376,2	543,0	393,7	445,0	-13,02
1998-2002	372,5	539,3	392,7	441,4	-12,38
1999-2003	368,3	533,7	395,1	435,5	-10,23
2000-2004	362,4	540,4	398,1	434,9	-9,25
2001-2005	359,4	530,8	396,1	430,6	-8,72
2002-2006	359,0	541,7	391,2	435,1	-11,23
2003-2007	369,5	544,0	395,7	442,3	-11,77
2004-2008	396,7	555,5	406,5	464,0	-14,16
priemer	379,0	550,3	395,4	449,9	-13,81

Rozdelené na zhluky podobného stavu a trendu vyzerá priestorová situácia plastickejšie. Na obrázku číslo 2 je časový priebeh relatívnej úmrtnosti na vybrané diagnózy („výber“) v štyroch podobách: ako prvý (lepší) zhluk podľa fuzzy c-zhlukovej analýzy, druhý (horší) zhluk, celoslovenský štandard a stav v južnej časti VÚC Trnava plus okrajové obce okolitých okresov.

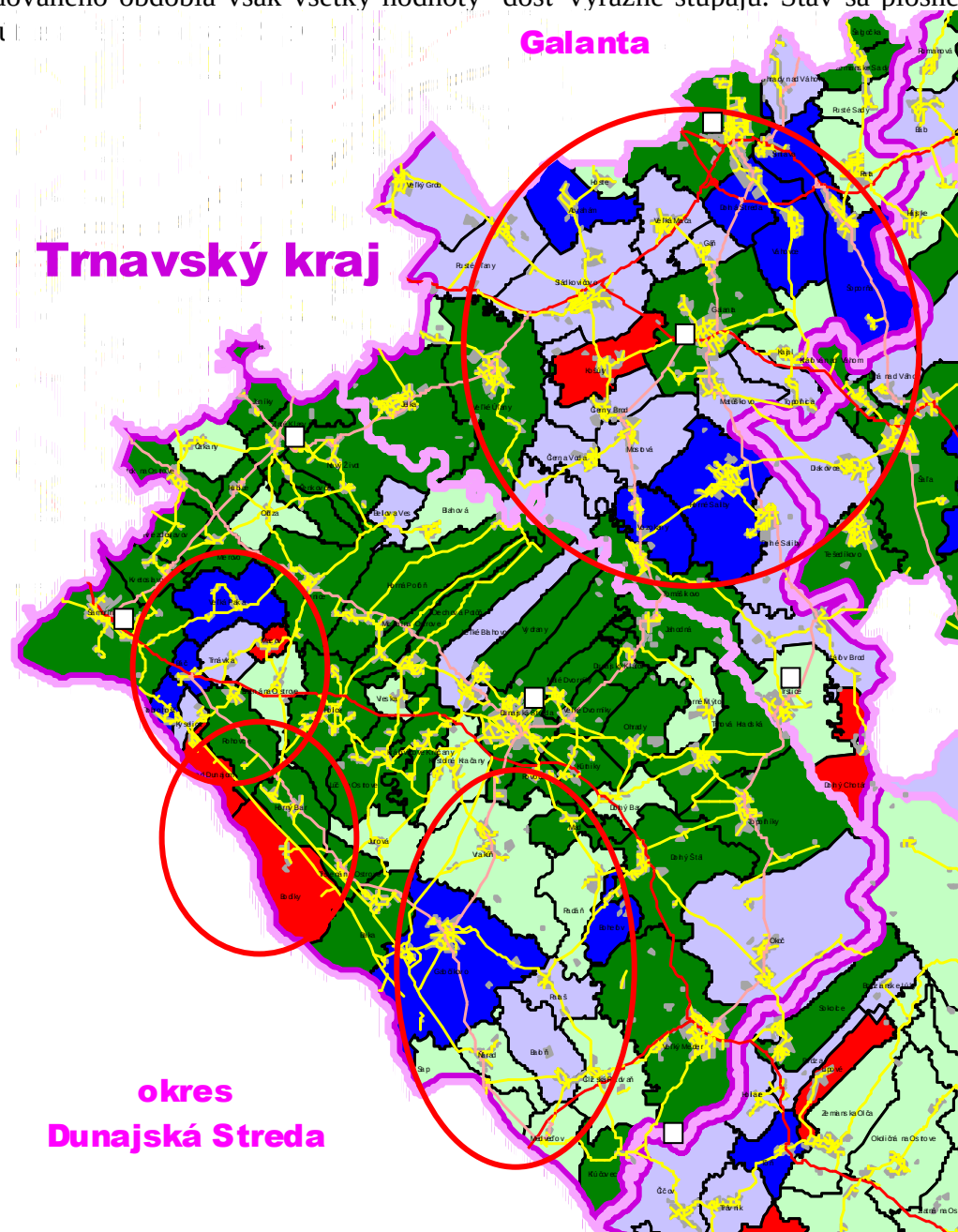


Obrázok 2: Relatívna úmrtnosť na vybrané príčiny 1994-2008

Prvý – lepší zhluk mal klesajúcu tendenciu a hneď v druhom časovom období rokov 1995-1999 klesol pod celoslovenskú hodnotu. Skoro po celý sledovaný čas sa aj držal pod Slovenskom, avšak v ostatných troch časových obdobiach začal stúpať podobne ako celé Slovensko, lenže oveľa strmšie. Ak bude tento trend dodržaný, je už dnes aj lepší južný zhluk nad celoslovenskými hodnotami.

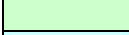
Horší zhluk JUH je o 40-50% vyšší ako lepší zhluk JUH a o 35-51% vyšší ako SR. Priemer oblasti je po celý čas horší, ako je Slovensko a to priemerne o 13%. Priemer oblasti je umiestnený blízko lepšieho zhluku, čo napovedá tomu, že dobrých obcí je v tejto analyzovanej skupine viac ako zlých.

Ku koncu sledovaného obdobia však všetky hodnoty dosť výrazne stúpajú. Stav sa plošne zhoršuje absol



Obrázok 3: Podrobnejší pohľad na okresy Dunajská Streda a Galanta, fuzzy c-zhluková analýza stavu a trendu úmrtnosti na vybrané diagnózy v rokoch 1994 až 2008

legenda k mape

	1. zhluh, ale lepší ako prvý
	1. zhluh - dobrý
	1.fuzzy, ale horší ako dobrý
	2. fuzzy, horší ako dobrý, ale lepší ako zlý
	2.zhluh - zlý
	2.fuzzy, ale horší ako zlý

biely štvorček na mape označuje stanicu ZS z rokov 2006-2009,
kruh (elipsa) označuje oblasti so zhoršeným stavom

Na mape obcí je výrazne vidieť, ako sú niektoré obce ťažko dosiahnuteľné v rámci existujúcej cestnej siete. Usudzujeme tak z toho, že zhluhy iných hodnôt ako dobrých sa výrazne ťahajú pozdĺž niektorých ciest. Podozrivé úseky sme označili elipsou.

Z mapy sa dá usúdiť, že dobrý zhluh a jeho v podstate nižšiu hodnotu ako Slovensko spôsobil najmä okres Dunajská Streda a Nitra, čiastočne aj okres Komárno. V tomto širšom susedskom porovnaní vychádza okres Galanta zle. Znovu pripomínáme, že okresy, ktoré sú v tejto analýze, majú podobné prírodné podmienky a nevidíme žiadny dôvod, prečo by mali byť rozdielne. Ak je v takomto susedstve jeden okres horší, je to skutočný problém, ktorý je potrebné riešiť.

Okres Dunajská Streda má veľmi veľký problém v obciach za Dunajom. Opakovane sa ukazuje ťažká dostupnosť obcí za prírodnou prekážkou. Ide o obce Vojka nad Dunajom a Bodíky.

Ostatné horšie oblasti sa asi ťažko lokalizujú nejakým výrazným bodom. Ako typicky zlé vychádza Gabčíkovo a potom celá široká oblasť smerom ku Komárnu je fuzzy – nejasná, niekedy skôr zlá, niekedy skôr dobrá. Ďalšia zhoršená oblasť je skupinka obcí s centrom v Macove, ktorý je zaradený ako horší ako zlý (červený). Podobné sú s ním susediace obce Veľká Paka, Trnávka, ďalej Kyselica, Báč a Dobrohošť. Tieto obce spája možno iba to, že sú ďaleko odvádiť. Celkovo však okres Dunajská Streda, až na spomínané obce za Dunajom, nepôsobí zle.

Iný je okres Galanta, ktorý sa ocitol celý v červenej elipse ako podozrivý. Tu nevieme stanoviť žiadnu čiastkovú horšiu oblasť. Je v nej celý okres okrem okresného mesta Galanta, ďalšieho mesta Sereď a dvoch veľkých obcí Veľké Úľany a Tomášikovo. Ako horšia ako zlá (červené) sa javí obec veľká Košúty. Podľa informácií, ktoré má táto obec zverejnené na internete, nie je tu žiadny špeciálny ústav, ktorý by mohol zhoršovať štatistiky obce. Samotná obec má cca 1500 obyvateľov, nejde o žiadnu hru malých čísel. Asi by stála za bližší rozbor. Ako aj celý okres Galanta.

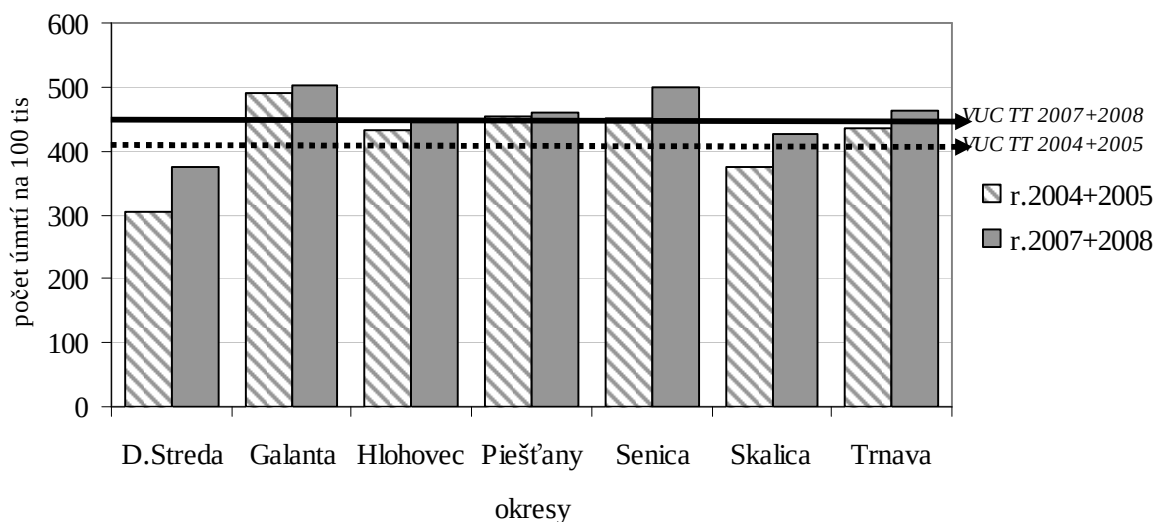
Jasným problémom sú prírodné a umelé prekážky. V prípade juhu ide o Dunaj a o obce, ktoré sú položené za ním. Inak sú problémy sporadické, roztrúsené, nevidíme dôvod, pre ktorý k nim prichádza. V okrese Dunajská Streda sme našli okrem zadunajských obcí dve súvislé väčšie problematické lokality. V okrese Galanta sme naopak našli iba štyri dobré obce v rámci celého zlého okresu.

Opakujeme, že prinajmenšom v obciach, ktoré sú vyfarbené červenou farbou, by sa malo uvažovať o núdzovom posilnení zdravotnej pohotovosti ešte iným spôsobom, napríklad spoluprácou s lokálnym lekárom, ktorý by dostal informáciu zároveň s miestopríslušnou stanicou Záchrannej zdravotnej služby a než táto dorazí na miesto, urobil by kvalifikovanú a najmä včasnú prvú pomoc.

4.3. Súhrnné hodnotenie všetkých oblastí kraja Trnava

Po zhodnotení všetkých troch oblastí kraja sme urobili jednoduché porovnanie dvoch rokov pred zavedením nového systému záchrannej zdravotnej služby v roku 2006 a zatiaľ dva ukončené po jej zavedení. Nový systém sa nazýva KOS (Krajské operačné stredisko) [6].

Výsledok nie je povzbudzujúci. Zdá sa, že nový systém prinajmenšom nie je lepší ako starý, ešte sa nevžil a neprináša očakávané výsledky, alebo jednoducho nie je dobrý.



Obrázok 4: porovnanie relatívnej úmrtnosti na vybrané diagnózy v okresoch VÚC Trnava pred a po zavedení nového systému zdravotnej záchranej služby (KOS) v roku 2006

V celom samosprávnom kraji, ale aj vo všetkých okresoch bez výnimky je badateľné zvýšenie vybranej úmrtnosti, pričom by sa očakávala prinajmenšom stagnácia, pokiaľ nie priamo jej zníženie. V ďalších výpočtoch sme zistili, že v okrese Dunajská Streda, kde je najvyšší rozdiel medzi vybranou mortalitou v rokoch 2004+2005 a 2007+2008, sa preukazuje zvýšila dokonca aj hrubá úmrtnosť, čo už je skutočne čo povedať pri takom robustnom indikátore.

Celé Slovensko má stúpajúci trend v úmrtnosti na vybrané diagnózy, ktorý v ostatných rokoch stúpa až dramaticky. Nie je to iba špecialita Trnavského kraja. Medziročný prírastok v posledných troch rokoch činí 10% ako na Slovensku, tak v kraji Trnava.

Je preukazná korelácia medzi podielom staršieho obyvateľstva a hodnotou indikátora **výber** (úmrtnosť na vybrané diagnózy). Podľa regresnej rovnice sme vypočítali modelové hodnoty pre všetky obce, ktoré zohľadňujú podiel starších občanov vo veku 65 a viac rokov. Toto modelové vyrovnanie sa nepremietlo do podstaty výsledkov. Obce, ktoré mali vysoké hodnoty indikátora **výber**, mali ich neprimerane vysoké aj po vekovom prepočte. Vek nie je jediným rozhodujúcim činiteľom v tejto úmrtnosti.

Prírodné prekážky ako pohoria alebo rieky, ale aj diaľnica, sú limitujúcim faktorom, ktorý výrazne zvyšuje hodnotu indikátora **výber** v obciach, ktoré ležia „za“ z pohľadu okresného mesta alebo inej ambulancie záchranej služby. Čo sa týka hôr, tam je jedno kde v rámci okresu sú, zlá situácia je všade.

Podobne sú pravdepodobne nedostatočne obslužené obce, ktoré ležia „v kútoch“ okresu. Najhoršie sú spravidla tie najvzdialenejšie, na hranici okresu. Podobná situácia sa opakuje na hranici susediaceho okresu a tak vznikajú celkom nelogické súvislé oblasti horšieho stavu, ktorý sa nedá narýchlo vysvetliť inak, iba že tam je odvšadiaľ ďaleko.

Obce, ktoré majú sociálne ústavy, napríklad Domovy dôchodcov, alebo niektorej inej sociálnej starostlivosti, sú výrazne skutočne veľmi zlé v tomto parametre (aj iných). Ich zhoršenie je v ráde stoviek percent od obvyklého stavu aj po zohľadnení veku obyvateľstva.

Porovnanie miest a dedín dopadlo ako výrazný rozdiel v neprospech vidieka. Mestá majú všeobecne dobré hodnoty tohto indikátora, 28 miest z 31 bolo zaradených do prvého – vyhranene dobrého zhluku.

Pri porovnaní dvojročníc pred zavedením nového systému záchranej služby a dvojročnice po jeho zavedení, s vynechaním roka, kedy sa zavádzal (2006) sme zistili štatisticky preukazné zhoršenie indikátora **výber**. Toto zhoršenie bolo štatisticky preukazné

ako v rámci celého VÚC Trnava, tak samostatne vypočítané aj v rámci okresu Dunajská Streda. V okrese Dunajská Streda sa v sledovanom období aj preukázne zvýšila hrubá úmrtnosť obyvateľstva.

5. Záver

Návrhy a odporúčania

Navrhujeme, aby boli podrobnejšie preskúmané vytypované oblasti, ktoré sa nám javia ako horšie v rámci zhlukovej analýzy.

Ďalej aby obce, ktoré sú označené červenou farbou na mape a boli zaradené do fuzzy zhluku „horšie ako zlé“, dostali okamžité posilnenie záchranej služby napríklad zmluvou s miestnym lekárom (bežný nemecký systém tzv. Notarzt).

Ďalší návrh tkvie v preskúmaní činnosti sociálnych ústavov a vo vypracovaní plánu na ich zlepšenie. Nielen v Trnave, na celom Slovensku sa obce s ústavmi javia nepomerne horšie aj pri všetkých vekových štandardizáciách.

Celkovo je potrebné posilniť zdravotnú organizáciu na vidieku, ktorá výrazne zaostáva za mestom. Všetky mestá sa javili ako „dobrý zhluk“, väčšina vidieka ako „zlý zhluk“.

Je potrebné, aby jednotlivé okresy lepšie spolupracovali a spoločne riešili nedostatočné „kúty“, t.j. obce, ktoré boli zaradené v horšom zhluku na oboch stranách spoločných hraníc okresov bez žiadneho iného dôvodu, iba že tam je ďaleko od ambulancie RZS.

Je možné, aby sa Trnavský samosprávny kraj zlepšil, má na to rezervy. Ako celok sa javí ako mierne horší ako jeho bezprostrední susedia. V rámci okresov vidíme stabilne dobrý a stále zlepšujúci sa okres Skalica a síce v absolútnom prevedení ešte stále dobrý, ale razantne sa zhoršujúci okres Dunajská Streda, ktorý potrebuje skutočne rýchle zhodnotenie tohto javu.

6. Literatúra

- [1] LETKOVIČOVÁ, M., 2009. Priestorová štúdia vybranej úmrtnosti v rokoch 1994-2008 v Trnavskom samosprávnom kraji: Environment, a.s., Nitra.
- [2] ŠTATISTICKÝ ÚRAD SR 1995 – 2009: Veková štruktúra obcí v rokoch 1994 až 2008, Jednotlivé záznamy o úmrtí v rokoch 1994 až 2008.
- [3] STEHLÍKOVÁ, B., 2004. Nový metodologický prístup k prognózovaniu demografických časových radov: Acta oeconomica et informatica, roč.7, č.2 (2004), s.49-52.
- [4] ELLIOT, P.-WAKEFIELD, J. - BEST, N.- BRIGGS, D., 2001. Spatial Epidemiology, Methods and Applications: OXFORD UNIVERSITY PRESS, ISBN 0-19-851532
- [5] MEDZINÁRODNÁ ŠTATISTICKÁ KLASIFIKÁCIA CHORÔB A PRÍBUZNÝCH ZDRAVOTNÝCH PROBLÉMOV. 10.revízia, WHO GENEVA 1992, ISBN 80-215-0282-7
- [6] Operačné stredisko záchranej zdravotnej služby. Krajské operačné stredisko ZZS Trnava. http://www.oszss-sr.sk/buxus/docs/Opravenie_KOS_TT.doc

Adresa autora:

Mária Letkovičová, Ing., CSc.
Jaskyňová 3
949 01 Nitra
letkovicova@environment.sk

Hana Zach, Ing., PhD
Čajkovského 11
949 11 Nitra
zach_hana@yahoo.co.uk

2x2 Kontingenčné tabuľky a diagnostické testy

2x2 Contingency tables and diagnostic tests

Ján Luha

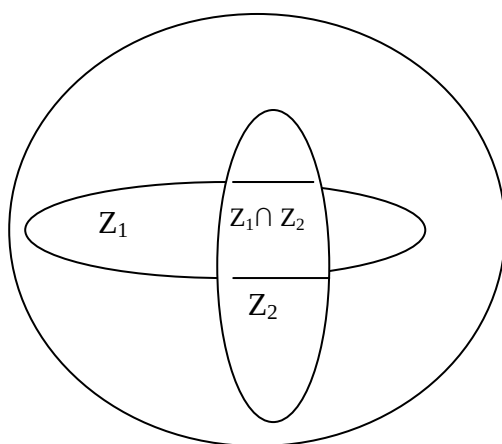
Abstract: In this article we present use of 2x2 Contingency tables in diagnostic tests in medicine. Illustrations are made on concrete examples.

Key words: 2x2 contingency tables, diagnostic tests, sensitivity, specificity, ROC curve.

Kľúčové slová: 2x2 kontingenčné tabuľky, diagnostické testy, senzitivita, špecificita, ROC krivka.

1. Úvod

Vzťah dvoch javov reprezentovaných znakmi Z_1 a Z_2 môžeme vyjadriť pomocou 2x2 kontingenčnej tabuľky ale aj množinovo či graficky. Javy majú rôzne schémy zvislostí, ktoré vidno z rozličných možností vzájomných prienikov. Uvedieme malú ilustráciu. V priestore Ω máme javy Z_1 a Z_2 s prienikom $Z_1 \cap Z_2$. Množinové vyjadrenie obvykle znázorňujeme napríklad obrázkom:



Kvôli názornosti možno použiť zjednodušenie, ktoré zodpovedá príslušným kontingenčným tabuľkám, pričom zohľadňuje veľkosti prienikov, rovnako ako sú početnosti v kontingenčných tabuľkách, teda napríklad:

	Z_2	Z_2'
Z_1	$Z_1 \cap Z_2$	$Z_1 \cap Z_2'$
Z_1'	$Z_1' \cap Z_2$	$Z_1' \cap Z_2'$

Kde $Z_i' = \Omega - Z_i$, $i = 1, 2$. Pričom Z_i označuje prítomnosť javu i a zase Z_i' neprítomnosť daného javu, pre $i=1, 2$.

2. 2x2 kontingenčné tabuľky

2x2 kontingenčné tabuľky majú široké použitie pri overovaní rozličných hypotéz. Kontingenčná tabuľka zaznamenáva početnosť spoločného výskytu dvoch binárnych znakov Z_1, Z_2 , možno ju zostaviť v tvare:

Z_1 / Z_2	Z_2	$\Omega - Z_2$	Spolu
Z_1	n_{11}	n_{12}	$n_{1.}$
$\Omega - Z_1$	n_{21}	n_{22}	$n_{2.}$
Spolu	$n_{.1}$	$n_{.2}$	n

Kde $n_{2.} = n - n_{1.}$, $n_{.2} = n - n_{.1}$, $n = n_{11} + n_{12} + n_{21} + n_{22} = n_{1.} + n_{2.} = n_{.1} + n_{.2}$.

Zrejme tiež platí: $n_{12} = n_{1.} - n_{11}$; $n_{21} = n_{.1} - n_{11}$; $n_{22} = n - n_{1.} - n_{.1} + n_{11}$.

Relatívne početnosti, ktoré sú odhadom príslušných pravdepodobností, možno tiež vyjadriť, podobne ako absolútne početnosti, v tvare kontingenčnej tabuľky:

Z_1 / Z_2	Z_2	$\Omega - Z_2$	Spolu
Z_1	p_{11}	p_{12}	$p_{1.}$
$\Omega - Z_1$	p_{21}	p_{22}	$p_{2.}$
Spolu	$p_{.1}$	$p_{.2}$	1

Vzťahy pre relatívne početnosti 2x2 kontingenčnej tabuľky sú zrejmé. Platí: $p_{2.} = 1 - p_{1.}$, $p_{.2} = 1 - p_{.1}$, $1 = p_{11} + p_{12} + p_{21} + p_{22} = p_{1.} + p_{2.} = p_{.1} + p_{.2}$, a tiež: $p_{12} = p_{1.} - p_{11}$; $p_{21} = p_{.1} - p_{11}$; $p_{22} = 1 - p_{1.} - p_{.1} + p_{11}$.

Za predpokladu, že marginálne početnosti oboch znakov ($n_{1.}$, $n_{2.}$ a $n_{.1}$, $n_{.2}$) sú fixné a tým aj rozsah výberu n , pre náhodný výber pravdepodobnosť kontingenčnej tabuľky vypočítame pomocou exaktného vzťahu:

$$P((n_{11}, n_{12}, n_{21}, n_{22})) = \frac{(n_{1.}! n_{2.}! n_{.1}! n_{.2}!)}{(n! n_{11}! n_{12}! n_{21}! n_{22}!)}.$$

Za uvedeného predpokladu fixných marginálnych početností oboch znakov môžeme zostrojiť **priestor všetkých možných kontingenčných tabuliek** a vypočítať ich pravdepodobnosti. Ľahko môžeme vypočítať počet všetkých možných kontingenčných tabuliek pre dané marginálne početnosti $n_{1.}, n_{2.}, n_{.1}, n_{.2}$. Nech kontingenčná tabuľka je usporiadaná tak, že $n_{1.}$ je najmenšia marginálna početnosť, potom počet všetkých kontingenčných tabuliek priestoru určeného danými marginálnymi početnosťami je $n_{1.} + 1$. Keď ďalej predpokladáme, že $n_{.1}$ je menšia alebo rovná $n_{2.}$, tak priestor všetkých kontingenčných tabuliek je daný početnosťami $n_{11}, n_{12}, n_{21}, n_{22}$ takto:

$$n_{11} = i, n_{12} = n_{1.} - i; n_{21} = n_{.1} - i, n_{22} = n - n_{1.} - n_{.1} + i, \text{ pre } i = 0, 1, \dots, n_{1.} + 1.$$

Príklad 1: Uvažujme kontingenčné tabuľky s fixnými marginálnymi početnosťami $n_{1.}, n_{2.}, n_{.1}, n_{.2} = 4, 4, 4, 4$, teda $n=8$. Potom priestor všetkých kontingenčných tabuliek s rovnakými marginálnymi početnosťami je 5 tabuliek:

i	0	1	2	3	4																				
	<table><tr><td>0</td><td>4</td></tr><tr><td>4</td><td>0</td></tr></table>	0	4	4	0	<table><tr><td>1</td><td>3</td></tr><tr><td>3</td><td>1</td></tr></table>	1	3	3	1	<table><tr><td>2</td><td>2</td></tr><tr><td>2</td><td>2</td></tr></table>	2	2	2	2	<table><tr><td>3</td><td>1</td></tr><tr><td>1</td><td>3</td></tr></table>	3	1	1	3	<table><tr><td>4</td><td>0</td></tr><tr><td>0</td><td>4</td></tr></table>	4	0	0	4
0	4																								
4	0																								
1	3																								
3	1																								
2	2																								
2	2																								
3	1																								
1	3																								
4	0																								
0	4																								

Pravdepodobnosti jednotlivých tabuliek:

P(i)	0,0143	0,2286	0,5143	0,2286	0,0143
-------------	--------	--------	--------	--------	--------

P-hodnoty dvojstranného a jednostranného Fisherovho exaktného testu nezávislosti tabuliek:

P₁(i)	0,0143	0,2429	0,7571	0,2429	0,0143
P₂(i)	0,0286	0,4857	1,0000	0,4857	0,0286

Korelačné koeficienty javov príslušných kontingenčných tabuliek:

R(i)	-1	-0,5	0	0,5	1
-------------	----	------	---	-----	---

Vidno, že iba tabuľky pre $i=0$ a 4 sú „signifikantné“, čo znamená závislosť dvoch znakov, resp. odlišnosť podielov, keď komparujeme výsledky za podsúbory jedného znaku voči druhému. Každá iná tabuľka z tohoto priestoru ukazuje nezávislé znaky, resp. pri komparácii podielov ich zhodu.

Plastickejšie vyjadruje priestor kontingenčných tabuliek ďalší príklad.

Príklad 2: Uvažujme kontingenčné tabuľky s fixnými marginálnymi početnosťami $n_{1.}, n_{2.}, n_{.1}, n_{.2} = 7, 15, 8, 14$, teda $n=22$. Potom priestor všetkých kontingenčných tabuliek s rovnakými marginálnymi početnosťami je 8 tabuliek. Okrem tabuliek uvádzame aj ich pravdepodobnosti, P-hodnoty jednostranného a dvojstranného Fisherovho exaktného testu a korelačný koeficient.

Tento priestor kontingenčných tabuliek obsahuje viac možností pre signifikantné výsledky.

i	0		1		2		3		4		5		6		7	
	0	7	1	6	2	5	3	4	4	3	5	2	6	1	7	0
	8	7	7	8	6	9	5	10	4	11	3	12	2	13	1	14
P(i)	0,0201		0,1409		0,3287		0,3287		0,1494		0,0299		0,0023		0,0000	
P_1(i)	0,0201		0,1610		0,4897		0,5103		0,1816		0,0322		0,0023		0,0000	
P_2(i)	0,0225		0,1932		1,0000		1,0000		0,3426		0,0524		0,0023		0,0000	
R(i)	-0,516		-0,314		-0,111		0,092		0,295		0,498		0,7008		0,9037	

O spoločnom rozdelení početností dvoch znakov môžeme vytvoriť viac hypotéz. Uvedieme najčastejšie z nich. Základná je o nezávislosti dvojice znakov, ktorá je ekvivalentná hypotéze homogenity – resp. rovnakého rozdelenia početností za podsúbory.

Hypotéza nezávislosti $H_0: p_{ij}=p_{i.}p_{.j}$, $i,j=1,2$. Pre 2×2 kontingenčné tabuľky je túto hypotézu možno vyjadriť vzťahom: $H_0: p_{11}=p_{1.}p_{.1}$,

alternatívne hypotézy: $H_1: p_{11} \neq p_{1.}p_{.1}$, $H_1: p_{11} > p_{1.}p_{.1}$, $H_1: p_{11} < p_{1.}p_{.1}$.

Testovacia štatistika: χ^2 , Fisherov exaktný test.

Hypotéza homogeneity $H_0: p_{11}=p_{21}$ a $p_{12}=p_{22}$,

alternatívne hypotézy: H_1 : opak H_0 .

Testovacia štatistika: χ^2 , Fisherov exaktný test.

Hypotéza o zhode marginálnych početností $H_0: p_{1.}=p_{.1}$, resp. $H_0: p_{12}=p_{21}$,

alternatívne hypotézy: $H_1: p_{1.} \neq p_{.1}$, $H_1: p_{1.} > p_{.1}$, $H_1: p_{1.} < p_{.1}$.

Testovacia štatistika: McNemar.

Ak možno pre oba znaky považovať množinu hodnôt za usporiadanú, tak možno vysloviť nulovú hypotézu :

Hypotéza linearity H_0 : s rastom početností (poklesom) prvého znaku rastie početnosť (klesá) aj druhého znaku.

alternatívna hypotéza: H_1 : opak H_0 .

Testovacia štatistika: Linear-by-Linear Association.

$H_0: p_{11} > p_{22}$, $p_{1.} < p_{.1}$.

Uvedené hypotézy a tiež ďalšie overujeme pri aplikáciách 2x2 kontingenčných tabuliek v rozličných štúdiách.

Rozlišujeme **tri spôsoby získania výberových súborov na analýzu 2x2 kontingenčných tabuliek komparatívnych štúdií**.

I. spôsob – prierezové (cross-sectional) štúdie - náhodným výberom vyberieme pevne daný rozsah n osôb u ktorých sledujeme výskyt, resp. absenciu dvoch javov reprezentovaných znakmi Z_1, Z_2 .

Príklad: (prevzaté z práce Fleiss J.L., Levin B., Myunghee Ch.P. (2003)) **Z_1 je vek matky**, kategorizovaný takto 1= vek ≤ 20 rokov; 2= vek nad 20 rokov, **Z_2 je hmotnosť plodu pri narodení** - s kategóriami 1= hmotnosť ≤ 2500 gramov a 2= hmotnosť nad 2500 gramov. Uvažujme hypotetický náhodný výber o rozsahu $n=200$ pôrodov na určitej klinike. Hypotetický výsledok je v kontingenčnej tabuľke:

Z_1 / Z_2	Z_2	$\Omega - Z_2$	Spolu
Z_1	10	40	50
$\Omega - Z_1$	15	135	150
Spolu	25	175	200

Obyčajne vychádzame z predpokladu, že nielen rozsah n , ale aj obe marginálne početnosti sú fixné. Za tohoto predpokladu môžeme počítať exaktné pravdepodobnosti a aplikovať Fisherov exaktný test, prípadne iné testy ako napríklad χ^2 McNemarov test a i.

Rozlišujeme ešte ďalšie možnosti – fixný je rozsah výberu n a marginálna početnosť jedného znaku a keď je fixný iba rozsah výberu n a marginálne početnosti znakov nie. V praxi ale tieto prípady riešime aproximačne rovnako ako prípad s oboma fixnými marginálnymi početnosťami.

II. spôsob – prospektívne a retrospektívne (Prospective and Retrospective) štúdie – náhodným spôsobom vyberieme n_1 osôb z tých čo majú jav Z_1 a n_2 z tých čo nemajú jav Z_1 a sledujeme výskyt, resp. absenciu javu Z_2 .

Pričom pre prospektívne štúdie zostavíme dva nezávislé výbery a skúmame prítomnosť a absenciu skúmaného javu Z_2 , napríklad chorobu, ako príklad využijeme skúmanie vzťahu veku matky a hmotnosti plodu uvedeného hore. Zostavíme dva náhodné výbery každý o rozsahu 100. Výsledok je v kontingenčnej tabuľke.

Z_1 / Z_2	Z_2	$\Omega - Z_2$	Spolu
Z_1	20	80	100
$\Omega - Z_1$	10	90	100
Spolu	30	170	200

Zostávajú zachované proporcie javov ako v pôvodnom príklade. 20 % mladých matiek (jav Z_1) rodí deti s menšou pôrodnou hmotnosťou a 10 % matiek vo vyššom veku (jav opačný ku Z_1) rodí deti s menšou pôrodnou hmotnosťou.

Retrospektívne (case-control) štúdie sú logicky síce obdobné ale skúmajú javy z iného pohľadu. Skúmame výber osôb so sledovanou vlastnosťou a výber ako kontrolnú skupinu. Využijeme znovu hypotetický príklad, ktorý môžeme skúmať aj retrospektívne. Znovu zostavíme dva nezávislé výbery o rozsahu 100 avšak za jav Z_2 . Výsledkom je kontingenčná tabuľka, ktorá zachováva proporcie javov z pôvodnej tabuľky.

Z_1 / Z_2	Z_2	$\Omega - Z_2$	Spolu
Z_1	40	23	63
$\Omega - Z_1$	60	90	137
Spolu	100	77	200

Keď uvažujeme hypotézu nezávislosti alebo ekvivalentnú hypotézu homogenosti dostávame **pre rovnaké rozsahy výberov** potenciálne rozdielne výsledky testov podľa troch typov štúdií:

Pre prierezovú štúdiu je $\chi^2=3,429$, a P-hodnota Fisherovho exaktného testu $P=0,083$, teda výsledok nie je signifikantný, aj pre jednostrannú alternatívu, keď $P=0,058$.

Pre prospektívnu štúdiu je $\chi^2=3,922$, a P-hodnota Fisherovho exaktného testu $P=0,073$, teda výsledok nie je signifikantný, ale pre jednostrannú alternatívu je $P=0,037$.

Pre retrospektívnu štúdiu je $\chi^2=6,697$, a P-hodnota Fisherovho exaktného testu $P=0,015$, teda výsledok je signifikantný a pre jednostrannú alternatívu je $P=0,007$.

III. spôsob – znáhodnené klinické experimenty (Randomized Controlled Trials) – náhodným spôsobom vyberieme n_1 osôb liečných daným spôsobom a taktiež náhodne vyberieme n_2 osôb liečných iným spôsobom (znak Z_1), prípadne kontrolnú skupinu osôb a sledujeme výskyt alebo absenciu choroby – znak Z_2 . Špeciálnou aplikáciou sú diagnostické testy, ktoré bližšie skúmame v nasledovnej kapitole.

Vo všetkých prípadoch je základom 2x2 kontingenčná tabuľka. Líšia sa úlohami, ktoré riešime a podľa toho vyberáme aj príslušné testy.

3. 2x2 kontingenčné tabuľky a diagnostické testy

Špeciálnou aplikáciou 2x2 kontingenčných tabuliek a príslušných testov sú diagnostické testy. Pri diagnostických testoch je obvyklé nasledovné označovanie početností 2x2 kontingenčnej tabuľky:

Výsledok testu	Skutočnosť		
	+	-	Spolu
+	a	b	a+b
-	c	d	c+d
Spolu	a+c	b+d	n

Iné označenie komparácie výsledku testu a skutočnosti:

Výsledok testu - znak A	Skutočnosť - znak B		
	chorá osoba	zdravá osoba	Spolu
pozitívny	a true-positive	b false-positive	a+b
negatívny	c false-negative	d true-negative	c+d
Spolu	a+c	b+d	n

Správne sú teda dva výsledky: true-positive a true-negative – čiže správne určenie chorej osoby pozitívnym testom a správne určenie zdravej osoby negatívnym výsledkom testu.

Taktiež sú **dva nesprávne výsledky:** false-positive – nesprávny pozitívny výsledok testu pre zdravú osobu a false-negative – nesprávny negatívny výsledok testu pre chorú osobu.

Výsledok diagnostického testu môže byť (+)=pozitívny (označme ako jav A) a (-)=negatívny (označme ako komplement symbolom A'), pričom skutočnosť, teda prítomnosť diagnostikovanej choroby môže byť (+)=pozitívny výskyt (jav B) a (-)=negatívny, čiže osoba je zdravá (komplement javu B označujeme ako B'). Život je samozrejme zložitejší a nemusí byť vždy jednoznačný ani výsledok testu a ani skutočnosť.

Pri skúmaní najjednoduchšej schémy komparácií, ktorou je 2x2 kontingenčná tabuľka, v výhodou aplikujeme Bayesovu vetu, ktorá využíva podmienené pravdepodobnosti. Najjednoduchšia schéma však skrýva veľa úskalí. Najprv definujeme základné podmienené pravdepodobnosti pre 2x2 kontingenčnú tabuľku.

Podmienená pravdepodobnosť pozitívneho výsledku testu, za predpokladu, že osoba je chorá je $Sens = P(A/B) = a/(a+c)$, čo tiež označujeme ako **senzitivitu**. Ďalej $P(A/B')$ je pravdepodobnosť pozitívneho výsledku testu zdravej osoby.

Podmienená pravdepodobnosť komplementárnych javov $Spec = P(A'/B') = d/(b+d)$ nazývame **špecificita**. Čím je pravdepodobnosť $P(A/B')$ menšia, resp. ekvivalentne $P(A'/B')$ väčšia, tým je test viac špecifický.

Podmienená pravdepodobnosť $P(B/A)$ sa nazýva aj **pozitívna prediktívna hodnota** PPV. Platí: $PPV = P(B/A) = (P(A/B) \cdot P(B)) / P(A) = a/(a+b)$.

Podmienená pravdepodobnosť $P(B'/A')$ sa nazýva **negatívna prediktívna hodnota** a platí: $NPV = P(B'/A') = (P(A'/B') \cdot P(B')) / P(A') = (P(A'/B') \cdot (1-P(B))) / (1-P(A)) = d/(c+d)$.

Prevalencia choroby je $\Pi = P(B) = (a+c)/n$.

Z uvedeného môžeme vyjadriť vzťahy:

$$PPV = (Sens.\Pi) / (Sens.\Pi + (1-Spec).(1-\Pi)),$$

$$NPV = (Spec.(1-\Pi)) / (Spec.(1-\Pi) + (1-Sens).\Pi)$$

Prehľadne a s využitím obvyklého značenia sú pravdepodobnosti diagnostického testu v tabuľke (s obvyklými anglickými názvami):

False negative: $c/(a+c)$	Sensitivity: $a/(a+c)$	Predictive value for positive test: $a/(a+b)$
False positive: $b/(b+d)$	Specificity: $d/(b+d)$	Predictive value for negative test: $d/(c+d)$

Na ilustráciu využijeme príklad priestoru kontingenčných tabuliek z predošlej kapitoly.

i

0	1	2	3	4
0 4	1 3	2 2	3 1	4 0
4 0	3 1	2 2	1 3	0 4

Vypočítame preň senzitivitu, špecifickosť a prevalenciu:

Sens(i)	0,0	25,0	50,0	75,0	100,0
Spec(i)	0,0	25,0	50,0	75,0	100,0
$\Pi(i)$	50	50	50	50	50

Tento umelý príklad názorne ukazuje prípad rovnakej prevalencie 50% a rôzne hodnoty senzitivity a špecificity. Najhorší je z hľadiska diagnostického testu príklad kontingenčnej tabuľky pre $i=0$, kedy sa nepodarilo nič správne určiť a najlepší je výsledok pre $i=4$, keď je absolútna úspešnosť hypotetického testu.

Uvedieme ešte jeden príklad priestoru kontingenčných tabuliek, tak ako je v predošlej kapitole. Tento príklad pre rovnakú prevalenciu 36,4% uvádza viac potenciálnych výsledkov testov, pričom test pre $i=0$ predstavuje najhoršiu možnosť a test pre $i=7$ zase najlepší.

i	0		1		2		3		4		5		6		7	
	0	7	1	6	2	5	3	4	4	3	5	2	6	1	7	0
	8	7	7	8	6	9	5	10	4	11	3	12	2	13	1	14
Sens(i)	0,0		12,5		25,0		37,5		50,0		62,5		75,0		87,5	
Spec(i)	50,0		57,1		64,3		71,4		78,6		85,7		92,9		100,0	
Π(i)	36,4		36,4		36,4		36,4		36,4		36,4		36,4		36,4	
PPV(i)	0,00		14,29		28,57		42,86		57,14		71,43		85,71		100,00	
NPV(i)	46,67		53,33		60,00		66,67		73,33		80,00		86,67		93,33	

4. Záver

Diagnostické testy sú dôležitým nástrojom analýzy klinických experimentov a nielen ich. Pri malých rozsahoch výberu je priestor možných kontingenčných tabuliek, ako sme ukázali v 2 kapitole veľmi malý a tým je obmedzený aj priestor na signifikantné výsledky. Zaujímavým a „signifikantným“ pokračovaním analýzy diagnostických testov a 2x2 kontingenčných tabuliek je ROC krivka (Receiver Operating Characteristic curve). Tejto téme sa budeme venovať na niektorej budúcej konferencii MEDSTAT.

5. Literatúra

- [1] Bland M.: Introduction to Medical Statistics, 3rd Edition, Oxford University Press 2000.
- [2] Feinstein A. R.: Principles of Medical Statistics. Boca Raton London New York Washington, D.C.© 2002 by Chapman & Hall/CRC.
- [3] Fleiss J.L., Levin B ., Myunghee Ch.P. (2003): Statistical Methods for Rates and Proportions. 3ed. Wiley , New York 2003.
- [4] Greenberg R.S., Daniels S.R., Flanders W.D., Eley J.W., Boring J.R. III.: Medical Epidemiology. Fourth Edition, McGraw-Hill 2005.
- [5] Chajdiak, J.(2003): Štatistika jednoducho. Statis Bratislava 2003, ISBN 808565928-X.
- [6] Chajdiak J. (2005): Štatistické úlohy a ich riešenie v Exceli. STATIS, Bratislava, ISBN 80- 85659-39-5.
- [7] Chajdiak J.(2007): Štatistika v exceli 2007. Statis, Bratislava 2009, ISBN 978-80-85659-49-8.
- [8] Kendall M. G., Stuart A. (1973): The Advanced Theory of Statistics, Volume 2, Inference and Relationship. Charles Griffin. London. Ruský preklad: Statističeskije vyvody i syjazi. Nauka Moskva 1973.
- [9] Luha J. (1985): Testovanie štatistických hypotéz pri analýze súborov charakterizovaných kvalitatívnymi znakmi. Vydal Odbor Výskumu programov ČST a divákov v SR. Bratislava 1985.
- [10] Luha J.(2003): Skúmanie súboru kvalitatívnych dát. EKOMSTAT 2003. SŠDS Bratislava 2003.
- [11] Luha J.: Korelácia javov. FORUM STATISTICUM SLOVACUM 2/2009. SŠDS Bratislava 2009. ISSN 1336-7420.
- [12] Riffenburg R. H.: Statistics in Medicine, Second Edition. Academic Press 2005.
- [13] Rosner B.: Fundamentals of Biostatistics. Fifth Edition, Harvard University. Duxbury Press 2000.
- [14] Sheskin D. J.: *Handbook of* PARAMETRIC and NONPARAMETRIC STATISTICAL PROCEDURES. THIRD EDITION, CHAPMAN & HALL/CRC 2004.
- [15] Woodward M.: Epidemiology. Study design and data analysis. Second Edition. Chapman and Hall/CRC 2005.

Adresa autora:

Ján Luha, RNDr., CSc.
Ústav lekárskej biológie, genetiky a klinickej
genetiky LF UK a FNsP
Sasinkova 4, Bratislava
jan.luha@fmed.uniba.sk

Morfológia a karyotypy databázy Abortus Morphology and Karyotypes of database Abortus

Daniel Böhmer, Tatiana Braxatorisová, Ján Luha, Jana Malová, Vanda Repiská,
Ján Vojtaššák

Abstract: In the article we present basic attributes of Abortus database. This database contains results about morphology and karyotypes from miscarriages and stillborn fetuses from Bratislava region from September 1992 to November 2009, which were analyzed (gross morphology & cytogenetics) at the Institute of Medical Biology, Genetics and Clinical Genetics, Faculty of Medicine, Comenius University in Bratislava. Our main goals were to identify Congenital Anomalies and determine karyotypes of examined embryos and fetuses.

Key words: miscarriages, stillbirth, morphology, karyotype database, attributes.

Kľúčové slová: potraty, mŕtvo narodení, morfológia, karyotypy, databáza, charakteristiky.

1. Úvod

Databáza Abortus zaznamenáva výsledky vyšetrenia plodu po samovoľnom potrate. Eviduje charakteristiky morfológie pomocou kódového značenia a aj ich textové charakteristiky a podobne karyotypy, demografické charakteristiky (*keď ich bolo možné zistiť*) ako vek plodu (v týždňoch, prípadne aspoň trimester), pohlavie plodu, vek matky v rokoch.

Vyšetrovali sme materiál zo spontánnych potratov, z indukovaných potratov u zamŕknutých potratov a z mŕtvych plodov z bratislavských kliník. Všetky vzorky boli vyšetrované morfológicky (vonkajšou obhliadkou) a dlhodobo *in vitro* kultivované, s cieľom dosiahnuť dostatočné množstvo cytogeneticky hodnotiteľných c-metafáz.

Nakoľko samovoľné potraty a mŕtvo rodené plody, ako neúspešná zložka ľudskej reprodukcie, obsahujú vysokú koncentráciu porúch vývoja a tiež kumulujú vysoký podiel chromozómových aberácií, našim cieľom bolo poskytnutie informácie gynekológovi o tom, či išlo o perspektívnu graviditu, prípadne prispieť k objasneniu etiopatogenézy pozorovaných anomálií. Tieto poznatky gynekológovi umožňujú optimalizovať starostlivosť o tehotnú ženu, vrátane genetickej konzultácie, a tak zvýšiť pravdepodobnosť úspešnosti nasledujúceho tehotenstva.

Morfologické vyšetrenie

Pri morfológickom hodnotení sme vonkajšou obhliadkou zisťovali prítomnosť jednotlivých súčastí plodového vajca (amniová membrána, choriové klky), prítomnosť embrya alebo plodu, jeho častí alebo štruktúr, ktoré dokazujú ich založenie (žltkový vak, pupočník). Použili sme klasifikáciu podľa Fujikuru a Fantela (1, 2).

Samovoľné potraty sa zaraďujú do štyroch morfológických tried s jedenástimi podskupinami (tab. 1).

Tabuľka 1. Morfológické zatriedenie samovoľných potratov

I. trieda	F	nekompletná vzorka s deciduou a časťami trofoblastu, fragment plodového vajca
II. trieda	RSC	prasknuté plodové vajce bez embryonálneho materiálu s úponom pupočníka, amnionu alebo so žltkovým vakom
	RS	prasknuté plodové vajce bez štruktúr uvedených pri RSC
III. trieda	IES	nepoškodené plodové vajce bez embryonálnych štruktúr - anembryomola
IV. trieda	EN	kompletné embryo bez vonkajších malformácií
	FN	plod bez vonkajších malformácií
	DE	dezorganizované embryo – anomálna diferenciácia embrya alebo rudimentárne embryo
	EF	embryo s fokálnymi anomáliami orgánov alebo orgánových systémov
	FF	plod s fokálnymi anomáliami orgánov alebo orgánových systémov
	IE	nekompletné embryo, pri ktorom nie je možné diagnostikovať normalitu alebo patológiu
	IF	nekompletné plod, pri ktorom nie je možné diagnostikovať normalitu alebo patológiu

Tento systém používa jasné kritériá, opiera sa o makroskopické zisťovanie prítomnosti či neprítomnosti definovaných štruktúr a je najmenej citlivý k arteficiálnemu skresleniu druhotnými zmenami, v dôsledku intrauterinnej retencie.

Informatívne materiály obsahovali embryo alebo plod, prípadne ich súčasti a umožňovali do určitej miery posúdiť priebeh embryogenézy.

Neinformatívne alebo málo informatívne skupiny

- F, RSC, RS, UF, UE

Informatívne skupiny:

- **ťažké vývojové defekty** - IES, DE
- **vrodené vývojové chyby novorodencov** - EF a FF
- **normálne embryá a plody** – EN a FN

Medzi ťažké vývojové defekty patria preukázané a suspektné anembryomoly (IES) a cylindrické, amorfné, rudimentárne embryá zahrnuté v skupine DE. Malformácie pri embryách a plodoch zaradených do skupín EF a FF zodpovedajú vrodeným vývojovým chybám novorodencov. Už pri tomto vyšetrení môžeme odhaliť ťažké patologické stavy, ktoré genetika a následne pôrodníka informujú o príčine intrauterinného úmrtia. Problémom ostávajú neinformatívne skupiny, ktoré pre neúplnosť získaného materiálu nemôžeme bližšie morfológicky analyzovať. Tieto skupiny predstavujú až 71,8 % materiálov zasielaných z klinických pracovísk.

Cytogenetické vyšetrenie

Frekvencia chromozómových aberácií a ich typy sú odlišné pri samovoľných potratoch v I., II. trimestri tehotnosti, pri mŕtvo narodených deťoch a pri novorodencoch (Vogel a Motulsky, 1986). Pri samovoľných potratoch v I. trimestri tehotnosti sa stretávame s trizómiami autozómov až v 50 – 70%.

Niektoré chromozómové aberácie, napr. Turnerov syndróm, sa vyskytujú u novorodencov aj u samovoľných potratov. Iné, napr. triploidie, sú skoro vždy nezlučiteľné so životom a vedú k reprodukčnému neúspechu. Niektoré chromozómové aberácie, napr. trizómia chromozómu č. 16, boli výlučne pozorované len pri samovoľných potratoch.

Vysoký podiel chromozómových aberácií nachádzame pri samovoľných potratoch s vývojovými anomáliami plodu, s hydropickou degeneráciou choriových klkov alebo pri prázdnych neporušených plodových vajciach (Boué a Boué, 1970).

Taktiež priemerný vek matiek pri samovoľných potratoch s chromozómovými aberáciami je zvýšený, kým pri samovoľných potratoch s normálnym karyotypom sa nelíši od veku rodičiek v populácii (Carr, 1977). Pri niektorých chromozómových aberáciách je zrejmé, že vyšší aj nižší vek rodičov, hlavne matky, ovplyvňuje vznik aberácie v zárodočných bunkách (Kučerová, 1978). Riziko Downovho syndrómu – najčastejšia chromozómová aberácia u človeka, sa zvyšuje so stúpajúcim vekom rodičov a tiež u veľmi mladých matiek. Pri monozómii X bol naopak častejšie pozorovaný nižší vek rodičiek (Lawry et al, 1976, Dejmek et al, 1985).

Materiálom, použitým na cytogenetické vyšetrenie boli fetálne, embryonálne a extraembryonálne tkanivá (pupočník, amnion, fragment kože a svalu a choriové klky), ktoré boli všetky kultivované dlhodobo. Trofoblast choriových klkov bol kultivovaný aj krátkodobo (12 – 24 hod.), prípadne priamo spracovaný (3). Preparáty boli pripravené štandardným postupom z bunkovej suspenzie a farbené konvenčne a G-prúžkovaním (4, 5).

Databáza v súčasnosti obsahuje 3333 spracovaných prípadov, postupne je dopĺňaná o nové vyšetrenia.

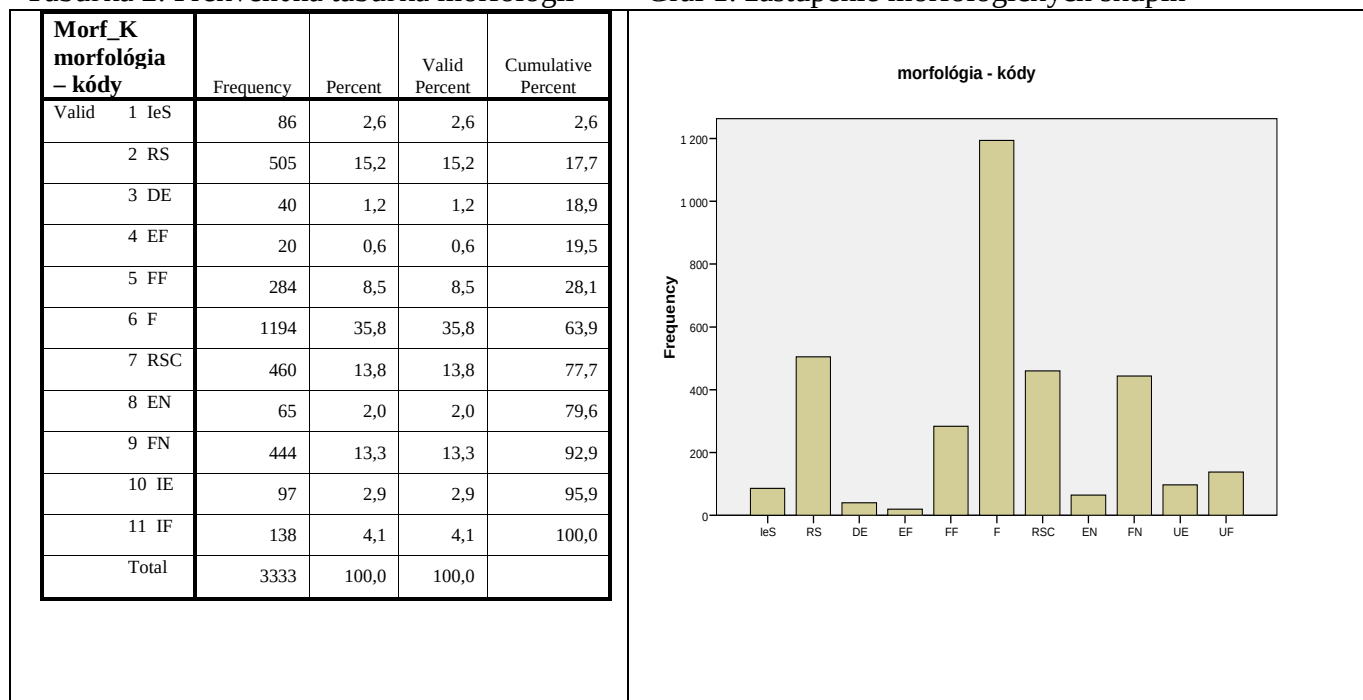
Pôvodne bola vedená v Exceli a kódy štatistických znakov boli alfanumerické, prípadne iba textové. Pre účely štatistickej analýzy boli alfanumerické kódy transformované na numerické. V budúcnosti bude spresnený záznam textových charakteristík morfológie a karyotypov.

2. Základné štatistické charakteristiky databázy Abortus

Charakterizáciu databázy Abortus, ktorá má v súčasnosti 3333 záznamov uvádzame vo frekvenčných tabuľkách a za numerické premenné, vek matky v rokoch a vývojový vek plodu v týždňoch, základné štatistické charakteristiky – čiže priemer a smerodajná odchýlka. Charakterizácia je doplnená o vybrané grafy, ktoré plastickejšie ilustrujú príslušné charakteristiky. Vzhľadom na obmedzený priestor na príspevok nekomentujeme zistené výsledky slovne. Tabuľky a grafy majú pre odborníka postačujúcu výpovednú hodnotu.

Tabuľka 2. Frekvenčná tabuľka morfológií

Graf 1. zastúpenie morfológických skupín



Na účely ďalších štatistických analýz sme vytvorili kategorizáciu znaku morfológia takto:

1 (neinf) = 2+6+7+10+11

2 (norm) = 8+9

3 (patol) = 1+3+4+6

Tabuľka 3. Frekvenčná tabuľka kategórií morfológií (**Morf_kateg**)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 neinf	2394	71,8	71,8	71,8
	2 norm	509	15,3	15,3	87,1
	3 patol	430	12,9	12,9	100,0
	Total	3333	100,0	100,0	

Až 71,8 % z celkového počtu tvoria neinformatívne skupiny F, RSC, RS, IF a IE. Z toho sme najčastejšie zachytili skupinu F (nekompletná vzorka s decíduou a časťami trofoblastu).

Ťažké vývojové defekty skupín IES a DE predstavujú 3,8 % vzoriek. Ide tu o veľmi včasnú poruchu embryonálneho vývoja. Tieto dve skupiny majú najnižší vývojový vek (2 - 3 týždne) a mimoriadne dlhé obdobie intrauterinnej retencie (5,5 – 7 týždňov).

Celkový podiel EF FF predstavuje 9,1 %, kým v novorodeneckej populácii je frekvencia vrodených vývojových chýb len 3,14 %.

Normálne embryá a plody – EN a FN – boli zastúpené len v 15,3 %.

Na ilustráciu textových záznamov uvádzame iba výbery, ktoré naznačujú potrebu ďalších úprav databázy. Je zrejmé, že neboli použité štandardizované záznamy a tak – bez ďalšej modifikácie – nie je možné textové záznamy využiť pri štatistickom spracovaní. To je ďalšia úloha pri transformovaní databázy.

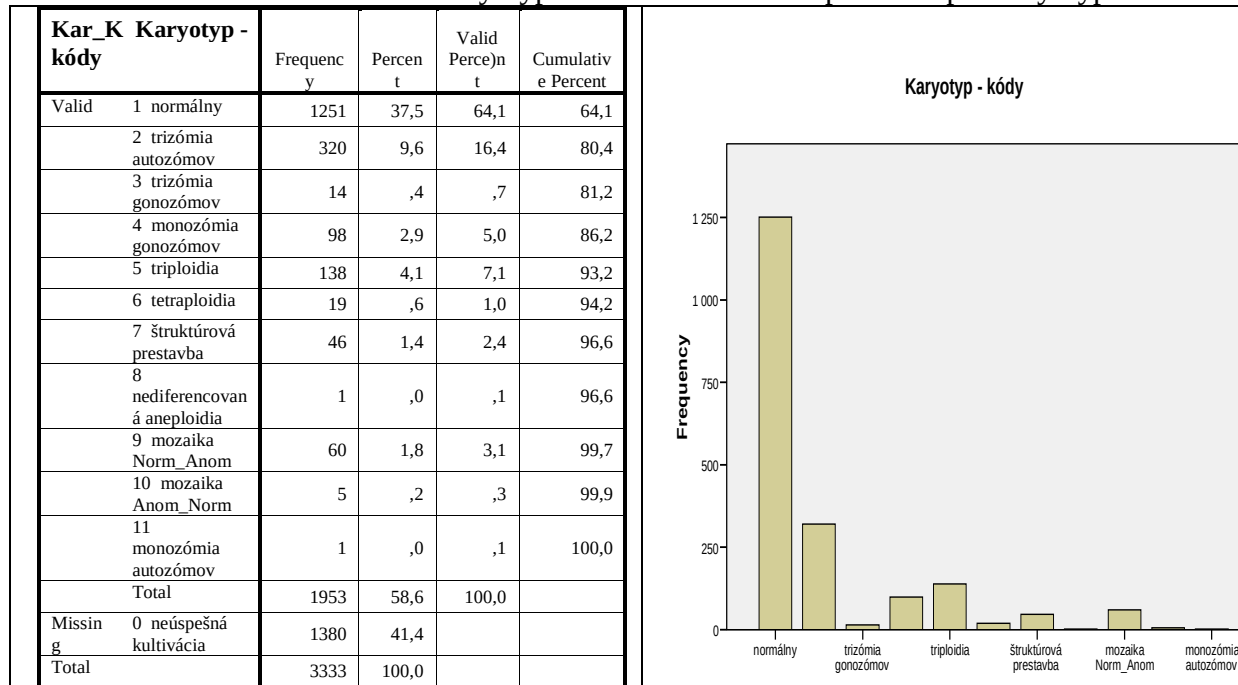
Tabuľka 4. Príklady slovných opisov morfológií (anomálií a malformácií) v databáze Abortus (**Morfológia – popisky**)

Morfológia
acranus, craniorachischisis po L5
acranus, cranioschisis
acranus, extrúzia orgánov dutiny brušnej, pes equinovarus bilat.
acranus, pravá orbita-anoftalmia, ľ.orbita-kryptoftalmia, hypoplastické nozdry, R III, malfomácia ušnic
acranus, spina bifida
anencefalus
anencefalus,
anencefalus, encefalokéla, spina bifida, omfalokéla, pupočník 2 cievy
anencefalus, R III.
anencefalus, spina bifida C-Th
anencefalus, susp. encefalokéla, R III, pes equinovarus LDK
anencephalus
M. Down
M.Down
M.Down,
MHC
pupočník - 2 cievy
pupočník s lumenmi 4 ciev,
pupočník-2 cievy
pupočník-2 cievy,

pupočník - 2 cievy
pupočník 2 cievy
pupočník s 2 cieva,i
pupočník s 2 cievami

Tabuľka 5. Frekvenčná tabuľka karyotypov

Graf 2. Zastúpenie skupín karyotypov



Na účely ďalších štatistických analýz sme vytvorili nasledovnú kategorizáciu znaku karyotyp (**Kar_K**):

0 (neúspešná kultivácia) = 0

1 (normálny) = 1

2 (iné) = 2-11

Čo sa týka cytogenetického vyšetrenia samovoľných potratov, spektrum chromozómových aberácií v tomto materiáli je oveľa bohatšie, v porovnaní s nálezmi u živonarodených detí. Pri samovoľných potratoch môžeme nájsť všetky numerické a štruktúrové aberácie, ktoré poznáme z karyotypov narodených detí, ale aj mnohé ďalšie, ktoré nie sú zlučiteľné s dosiahnutím vyšších vývojových štádií a s prežívaním po pôrode.

Najčastejšou skupinou chromozómových aberácií v databáze Abortus sú numerické aberácie – trizómie autozómov (9,6 %), monozómia X (2,9 %) a polyploidie, (triploidie – 4,1 % a tetraploidie 0,6%). Štruktúrové aberácie predstavujú 1,4 %.

Tabuľka 6. Frekvenčná tabuľka skupín karyotypov (**Kar_kateg**)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 normálny	1251	37,5	64,1	64,1
	2 iné	702	21,1	35,9	100,0
	Total	1953	58,6	100,0	
Missing	0 neúspešná kultivácia	1380	41,4		
Total		3333	100,0		

Podobne – ako textové spresnenia pri morfológii – aj pri textových charakteristikách karyotypov je pripravované rekódovanie opisov karyotypov, ktoré bude vhodnejšie na štatistické spracovanie. Na ilustráciu uvádzame príklady opisov.

Tabuľka 7. Príklady slovných opisov karyotypov

Karyotyp - opis	Karyotyp - opis
45, X, inv.17q	47, XX, -14, + t(11,14)
16, XX	48, XX, + 5, + 7
2n	48, XX, + 9, +16
3n	48, XX, + D(13), + 16
45, X - 46, XX	48, XXY, + D
45, X - 4n	48, XY, + 13, + 21
45, X	48, XY, + 2, + 21
46, XX	48, XY, + C(X), + 16
46, XY	48, XY, +2, +17
46, XY	48, XYY, + 17
47	48,XY,+8+14
47, XX ,1 C,5q+	49, XXXY
47, XX, -14, + t(11,14)	4n

Vývojový vek

Je charakterizovaný skutočným vekom plodu – v týždňoch, prípadne aspoň v trimestroch. Percentuálne zastúpenie podľa trimestrov je v tabuľke 8. Presnejšie zistený vek v týždňoch je zistený iba pre 1841 vzoriek z celkového počtu 3333, preto priemerný vek plodu v týždňoch 9,98 je iba informatívny.

Tabuľka 8. Frekvenčná tabuľka vývojového veku (v trimestroch) vyšetovaných vzoriek

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 I.trimester	2323	69,7	69,7	69,7
	2 II.trimester	935	28,1	28,1	97,7
	3 III.trimester	75	2,3	2,3	100,0
	Total	3333	100,0	100,0	

Tabuľka 9. Deskriptívna štatistika vývojového veku vyšetovaných vzoriek (v týždňoch)

	N	Minimum	Maximum	Mean	Std. Deviation
vek_plodu vek plodu v týždňoch	1841	1	34	9,98	5,973

Vek matky

Bol uvedený u 2231 vzoriek, z 3333 vyšetrených materiálov, čo je 67 %. Priemerný vek matky 28,77 roku je len odhad. Štruktúra veku matky podľa vekových kategórií, upravená na bázu zistených vekov ukazuje, že 30,7 % matiek bola vo vekovej kategórii 15-25 rokov, v kategórii 26-35 rokov bolo 55,7 % matiek a 13,6 % bolo vo veku 36-54 rokov. Výsledky veku plodu a veku matky sú znázornené v histograme.

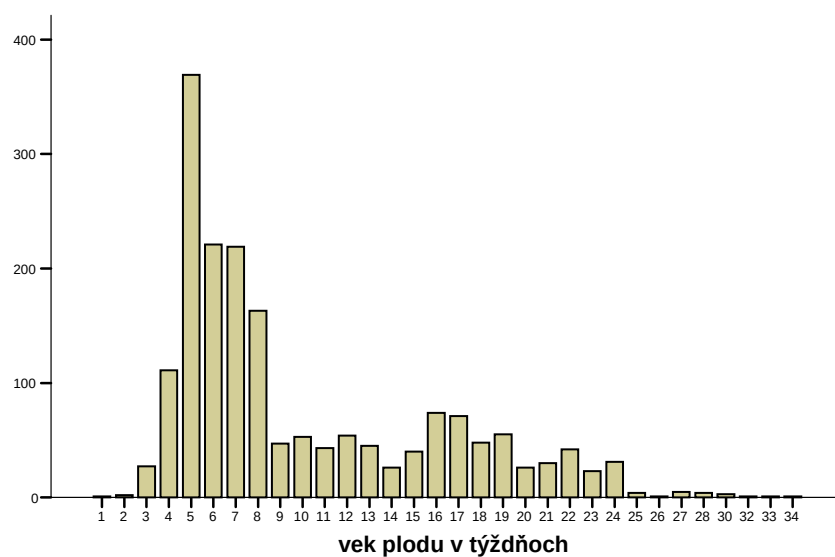
Tabuľka 10. Deskriptívna štatistika veku matky

	N	Minimum	Maximum	Mean	Std. Deviation
vek_matky	2231	15	54	28,77	5,992

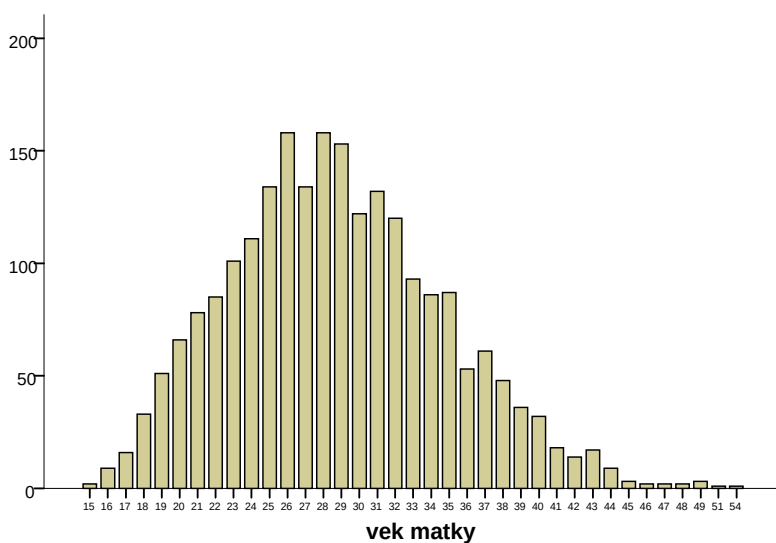
Tabuľka 11. Frekvenčná tabuľka veku matky – podľa skupín (vekM_kat)

		Frekvenc y	Percent	Valid Percent	Cumulati ve Percent
Valid	1 15-25	686	20,6	30,7	30,7
	2 26-35	1243	37,3	55,7	86,4
	3 36-54	303	9,1	13,6	100,0
	Total	2232	67,0	100,0	
Missing	System	1101	33,0		
Total		3333	100,0		

Graf 3. Histogram zastúpenia veku plodu (v týždňoch)



Graf 4. Histogram zastúpenia veku matky (v rokoch)



Pohlavie embrya plodu bolo možné zistiť len u 1914 vyšetrených vzoriek, čo je 57,4% z celkového počtu vzoriek. Mužské pohlavie bolo zistené v 43,2% prípadoch a ženské v 53,8% prípadoch.

Tabuľka 12. Frekvenčná tabuľka pohlavia

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 mužské	884	26,5	46,2	46,2
	2 ženské	1030	30,9	53,8	100,0
	Total	1914	57,4	100,0	
Missing	9 n	39	1,2		
	System	1380	41,4		
	Total	1419	42,6		
Total		3333	100,0		

3. Charakterizácia podľa úspešných kultivácií

Na doplnenie charakterizácie databázy sme vypočítali frekvenčné tabuľky za podsúbor úspešných kultivácií. Komparáciu s celým súborom necháme na čitateľa.

Tabuľka 13. Frekvenčné tabuľky sledovaných znakov za podsúbor úspešných kultivácií

Morf_kateg morfológia - kategorizácia					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 neinf	1337	68,5	68,5	68,5
	2 norm	357	18,3	18,3	86,7
	3 patol	259	13,3	13,3	100,0
	Total	1953	100,0	100,0	

Kar_kateg					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 normálny	1251	64,1	64,1	64,1
	2 trizómia autozómov	320	16,4	16,4	80,4
	3 trizómia gonozómov	14	,7	,7	81,2
	4 monoizómia gonozómov	98	5,0	5,0	86,2
	5 triploidia	138	7,1	7,1	93,2
	6 tetraploidia	19	1,0	1,0	94,2
	7 štruktúrová prestavba	46	2,4	2,4	96,6
	8 nediferencovaná aneuploidia	1	,1	,1	96,6
	9 mozaika Norm_Anom	60	3,1	3,1	99,7
	10 mozaika Anom_Norm	5	,3	,3	99,9
	11 monoizómia autozómov	1	,1	,1	100,0
	Total	1953	100,0	100,0	

Morf_K					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 IeS	44	2,3	2,3	2,3
	2 RS	263	13,5	13,5	15,7
	3 DE	28	1,4	1,4	17,2
	4 EF	10	,5	,5	17,7
	5 FF	177	9,1	9,1	26,7
	6 F	628	32,2	32,2	58,9
	7 RSC	302	15,5	15,5	74,3
	8 EN	34	1,7	1,7	76,1
	9 FN	323	16,5	16,5	92,6
	10 UE	62	3,2	3,2	95,8
	11 UF	82	4,2	4,2	100,0
	Total	1953	100,0	100,0	

Trimester		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 I.trimester	1254	64,2	64,2	64,2
	2 II.trimester	665	34,1	34,1	98,3
	3 III.trimester	34	1,7	1,7	100,0
	Total	1953	100,0	100,0	

pohlavie		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 mužské	884	45,3	45,3	45,3
	2 ženské	1030	52,7	52,7	98,0
	9 n	39	2,0	2,0	100,0
	Total	1953	100,0	100,0	

vekM_kat		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 15-25	621	31,8	33,0	33,0
	2 26-35	1000	51,2	53,1	86,1
	3 26-54	261	13,4	13,9	100,0
	Total	1882	96,4	100,0	
Missing	System	71	3,6		
Total		1953	100,0		

4. Závěry

Údaje registrované v databáze Abortus slúžia nielen na evidenciu charakteristík samovoľne potratených embryí a mŕtvo rodených plodov, ale aj na skúmanie závislostí sledovaných charakteristík, najmä morfológie a karyotypov a tiež na charakterizáciu podskupín podľa pohlavia, veku plodu, veku matky a ďalších súvislostí. Primárna funkcia databázy je evidovať stav, ale databázu možno využiť na štatistické analýzy ako – napríklad analýza závislostí, ktorá bude predmetom ďalších publikácií.

Databáza bude dopĺňovaná ďalšími prípadmi a postupne bude zavedený štandardizovaný záznam textových charakteristík morfológie a karyotypov.

5. Použitá literatúra

1. Boué, J., G., Boué, A.: Les aberrations chromosomiques dans les avortements spontanés humains. Presse med. 78, 14, 1970, 635-641
2. Böhmer D., Brašňová D., Luha J.: Špecifiká záznamu a spracovania dát z Národného registra pacientov s vrodenou vývojovou chybou. FORUM STATISTICUM SLOVACUM 7/2009. SŠDS Bratislava 2008. ISSN 1336-7420.
3. Bullerdiek, J., Bartnitzke, S., Schloot, W.: A rapid and simple sandwich – method used for chromosome analysis from small fetal and adult biopsy specimens. Clin Genet, 1979, 16, 433-437
4. Carr, D., H.: Chromosome studies in selected spontaneous abortions. III. Early pregnancy loss. Obstet. And Gynecol. 37, 5, 1971, 750-754
5. Dejmek, J., Vojtaššák, J., Izakovič, V., Šuchová, L., Malová, J., Paliderová, Z., Nováková, D.: Genetické aspekty spontánneho potratu vo vzťahu k vybraným etiologickým činiteľom. Záverečná správa, Bratislava VÚPL, 1985, 54
6. Fantel, A.G., Shepard, T.H., Vadheim – Roth, C. et al.: Embryonic and Fetal phenotypes: Prevalence and other associated factors in a large study of spontaneous abortion. In: Human Embryonic and Fetal Death. Porter, I.M., Hook, E.B., Eds., Acad Press, New York, 1988, 71
7. Fujikura, T., Froelich, L.A., Driscoll, S.G.: A Simplified Anatomic Classification of Abortions. Amer.J.Obstet.Gynecol., 95 (7), 1966, 902 – 905
8. Kučerová, M.: Cytogenetic analysis of human chromosomes and its value for the estimation of genetic risk. Mut. Res. 41, 1, 1978, 123-130

9. Lawry, R.B., Jones, D.C., Renwick, D.H.G., Trimble, B.K.: Down syndrome in British Columbia, 1972 – 1973: Incidence and Mean Maternal Age. *Teratology* 14, 1976, 29-34
10. Luha J.(2003): Matematickoštatistické aspekty spracovania dotazníkových výskumov. Štatistické metódy vo vedecko-výskumnej práci 2003, SŠDS, Bratislava 2003, ISBN 80-88946-32-8.
11. Luha J.: (2008) Metodologické aspekty zberu a záznamu dát otázok s možnosťou viac odpovedí. *FORUM STATISTICUM SLOVACUM* 7/2008. SŠDS Bratislava 2008. ISSN 1336-7420.
12. Seabright, M.: A rapid banding technique for human chromosomes. *Lancet* 2, 1971, p.971- 972
13. Simoni, G., Gimelli, G., Cuoco, C.: First Trimester Fetal karyotyping: One thousand diagnoses. *Hum. Genet.* 1986, 72, 203 – 209
14. Vogel, F., Motulsky, A.G.: *Human Genetics. Problems and Approaches.* Springer-Verlag, 1986, II.nd Ed, 80

Adresa autorov:

Ústav lekárskej biológie, genetiky a klinickej genetiky LF UK a FNŠP
Sasinkova 4, Bratislava

Daniel Böhmer, Doc., MUDr., PhD. daniel.bohmer@fmed.uniba.sk	Tatiana Braxatorisová, MUDr. tatiana.braxatorisova@fmed.uniba.sk
Ján Luha, RNDr., CSc. jan.luha@fmed.uniba.sk	Jana Malová, RNDr. jana.malova@fmed.uniba.sk
Vanda Repiská, Doc., MUDr., PhD. vanda.repiska@fmed.uniba.sk	Ján Vojtaššák, Prof., RNDr., PhD. jan.vojtassak@fmed.uniba.sk

Dlhodobá prognóza príjmov a výdavkov slovenského zdravotného systému s analýzou senzitivnosti

Long term forecast of the revenue and expense of the Slovak health care system and analysis his sensibility

Viliam Páleník, Vladimír Kvetan, Martin Mlýnek, Marek Radvanský

Abstract: Práca sa sústreďuje na otázku financovania slovenského zdravotného systému, ktoré je z národohospodárskeho pohľadu na svojej príjmovej stránke okrem iného závislé od makroekonomického vývoja a na výdavkovej stránke hlavne od demografického vývoja. Pri konštrukcii dlhodobej prognózy príjmov a výdavkov slovenského zdravotného systému boli zohľadnené aj ďalšie súvislosti a predpoklady ekonomického, legislatívneho a zdravotného charakteru. Ako nástroj modelovania bol zvolený cielene skonštruovaný model ekonometrického typu. Prognóza bola podrobená analýze senzitivnosti, ktorá ju umožnila relativizovať vzhľadom na možný alternatívny vývoj dôležitých makroekonomických a demografických indikátorov. Výsledky ukázali, že: jedno percentný dodatočný rast / pokles príjmov vedie k dodatočnému nárastu / poklesu príjmov zdravotného systému na úrovni 0,6 – 0,7 %; dodatočná jedno percentná zmena miery zamestnanosti vedie k nárastu / poklesu príjmov zdravotného systému o 0,05 % ; v prípade najpesimistickejšieho vývoja je možné očakávať deficit na úrovni 3 % z HDP ; deficitnosť dosiahne najpravdepodobnejšie úroveň deficitu 1,2 – 1,5 % z HDP v roku 2050. Ak by nedošlo ku zmene filozofie financovania zdravotného systému bude nutné voliť niektorú z týchto stratégií (resp. ich kombináciu): zvyšovať odvodové zaťaženie zamestnaných pracujúcich postupne na 21 % hrubej mzdy v roku 2050 ; rozšíriť odvodovú základňu ; zásadne zmeniť efektívnosť systému a zdravotné správanie obyvateľstva.

Key words: long term, forecast, Slovak health care system, revenue, expense, econometric model, sensibility analysis

Kľúčové slová: dlhodobá prognóza, slovenský zdravotný systém, príjmy, výdavky, ekonometrický model, analýzy senzitivnosti

1. Úvod

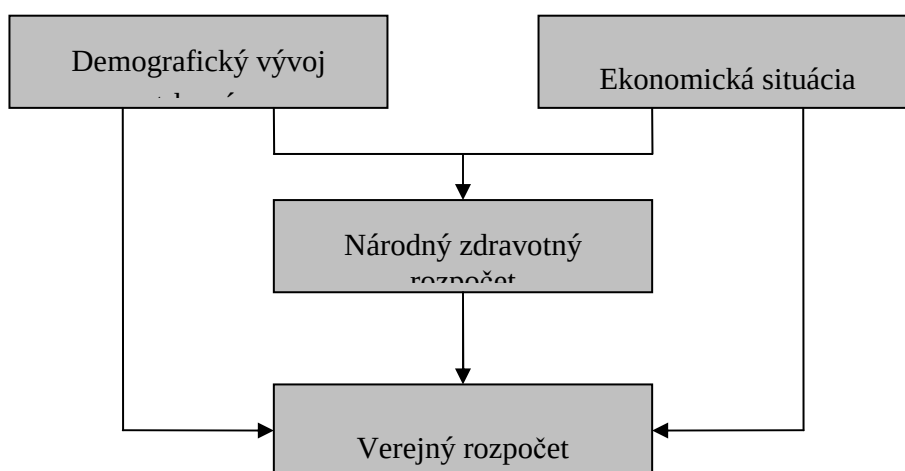
V roku 2004 sa Centrum pre európske politické štúdie CEPS z Belgicka ujalo vedenia konzorcia s 18 partnermi, ktorých väčšina bola členmi Európskej únie a Európskej siete hospodársko-politických výskumných inštitútov ENEPRI. Konzorcium bolo zostavené na riešenie projektu AHEAD – starnutie, zdravotný stav a determinanty zdravotných výdavkov. Tým bolo odštartované riešenie trojročnej výskumnej úlohy zameranej na budúci vývoj zdravotných výdavkov v rozšírenej Európskej únii.

Tento financovaný pod Šiestym výskumným rámcovým programom EÚ a zahŕňal penzum náročnej výskumnej práce, zahŕňajúcej množstvo workshopov, riešenie mnohých dátových a metodologických ťažkostí, organizačné problémy medzinárodného výskumného kolektívu, hľadanie jednotného chápania pojmov a metód medzinárodného interdisciplinárneho výskumného kolektívu, hlavne medzi jej ekonomickou a zdravotnou časťou. Riešiteľský kolektív z Ekonomického ústavu SAV sa do tejto práce aktívne zapojil.

Práce na projekte boli úspešne ukončené schválením záverečnej správy a uskutočnením záverečnej konferencie 28. a 29. júna 2007 v Bruseli. Medzičasom sa krajiny všetkých riešiteľských pracovísk stali členmi Európskej únie, vrátane Slovenska. Ide o jeden z prvých úspešne ukončených projektov európskych rámcových výskumných projektov zo socio-ekonomických vied s riešiteľskou účasťou slovenského výskumného pracoviska. Výsledky projektu dokázali, že starnutie a jeho ekonomické konsekvencie sú pre Slovensko vysoko aktuálne a ich riešenie je nutné intenzívne hľadať a realizovať. [1-4] Financovanie slovenského zdravotného systému je z makroekonomického pohľadu na svojej príjmovej stránke závislé od makroekonomického vývoja a na výdavkovej stránke hlavne od demografického vývoja. Pri konštrukcii dlhodobej prognózy príjmov a výdavkov nášho zdravotného systému je potrebné zohľadniť aj ďalšie súvislosti a predpoklady ekonomického, legislatívneho a zdravotného charakteru. Ako nástroj modelovania bol zvolený cielene skonštruovaný model ekonometrického typu. Prognóza bola podrobená analýze senzitivnosti, ktorá ju umožňuje relativizovať vzhľadom na možný alternatívny vývoj dôležitých makroekonomických a demografických indikátorov.

2. Charakteristika modelu

Model výdavkov systému zdravotného zabezpečenia pozostáva zo štyroch vzájomne prepojených modulov popisujúcich súvislosti demografického vývoja a trhu práce spolu s celkovou ekonomickou situáciou v SR a ich prepojenie na systém zdravotného zabezpečenia a verejný rozpočet (viď Schéma 1). Model je vystavaný na platforme MS Excel[®] ako sústava niekoľkých makrami navzájom prepojených zošitov.



S c h é m a 1: Základné väzby v modeli výdavkov systému zdravotného zabezpečenia

3. Modul trhu práce

Modul trhu práce je zameraný jednak na výpočet dopytu a ponuky práce a jednak na rozdelenie zamestnaných do jednotlivých typov zamestnaní.¹ Na základe demografickej prognózy v požadovanej vekovej štruktúre je v tomto module napočítaná najprv ponuka práce v jednotlivých vekových kategóriách pre dané pohlavie ako súčin počtu obyvateľov v danej skupine a miery ekonomickej aktivity. Celkový počet ekonomicky aktívnych obyvateľov je potom počítaný ako súčet ekonomicky aktívnych pre jednotlivé vekové skupiny a pohlavia.

$$ls_{ig} = n_{ig} * pr_{ig} \quad (3.1)$$

$$ls_{total} = \sum_i \sum_g ls_{ig} \quad (3.2)$$

ls - ponuka práce

n - počet obyvateľov v danej skupine

pr - miera ekonomickej aktivity

i - veková skupina

g - pohlavie

Dopyt po práci je v modeli počítaný na základe rastu oproti predošlému obdobiu, prípadne môže byť celkový dopyt po práci exogénne určený na základe makroekonomických modelov. V našom prípade sa používajú vstupy z modelov ISWExxqx. Celkový dopyt po práci je na základe pomerových ukazovateľov rozdelený na dopyt po práci pre jednotlivé pohlavia a vekové skupiny.

Na základe legislatívy a historických dát je následne počítaná zamestnanecká štruktúra (členenie zamestnaných osôb na zamestnancov, podnikateľov, atď.). Prognózované sú počty zamestnaných osôb podľa vekových a zamestnaneckých skupín, ako aj počet nezamestnaných osôb a miera nezamestnanosti v požadovanom horizonte prognózy.

4. Modul dopytu po zdravotnej starostlivosti²

Vstupnými dátami tohto modulu sú dáta popisujúce využívanie služieb zdravotnej starostlivosti jednotlivými vekovými skupinami obyvateľov. Dostupné údaje o priemernom počte návštev u lekára a priemernom počte prípadov hospitalizácie v špeciálnych vekových skupinách umožňujú aproximovať závislosť dopytu po nemocničnej resp. lekárskej starostlivosti od veku obyvateľa. Výstupom tohto modulu je teda aproximácia J-krivky³ pre Slovensko, ktorá pre konkrétnu prognózu vekovej štruktúry obyvateľstva umožňuje predikovať dopyt jednotlivých vekových kategórií po zdravotnej starostlivosti. Základná verzia modelu uvažuje s nemennou J-krivkou v ďalších obdobiach, ktorá je založená na historických dátach. Je však možné dopĺňať vlastné pravdepodobnosti využívania zdravotníckych služieb v jednotlivých vekových skupinách.

¹ V závislosti od platnej legislatívy vymedzujúcej jednotlivé skupiny prispievateľov do zdravotného systému.

² Staršia verzia modelu je popísaná v [5]. Aktuálne výsledky prognózy počítanej pomocou modelu ECM-ISWE05q1 sú v práci [6].

³ Označenie „J-krivka“ vystihuje relatívne nízky dopyt po službách zdravotnej starostlivosti zo strany obyvateľstva v strednom veku, zvýšený dopyt zo strany detí a enormné využívanie zdravotníckych služieb obyvateľmi dôchodkového veku.

$$HCU_{ig} = n_{ig} * p_{ig} \quad (3.3)$$

$$HCU_{total} = \sum_i \sum_g HCU_{ig} \quad (3.4)$$

HCU - využívanie zdravotníckych služieb

p - pravdepodobnosť využitia zdravotníckych služieb (z J-krivky)

n - počet obyvateľov v danej skupine

i - veková skupina

g - pohlavie

5. Modul zdravotného rozpočtu

Tento modul bilancuje príjmy a výdavky zdravotného systému. Výdavky celkového zdravotného systému sa počítajú ako súčet výdavkov na nemocničnú, lekársku a dlhodobú starostlivosť. Vývoj každej kategórie výdavkov pritom závisí od dopytu po zdravotnej starostlivosti napočítaného v module dopytu po zdravotnej starostlivosti a od jednotkových nákladov, ktorých tempo rastu je jedným z exogénnych parametrov modelu. Jednotkové náklady môžu byť stanovené pre všetky vekové kategórie a obe pohlavia rovnako, prípadne štruktúra modelu umožňuje individuálne nastavenia jednotkových nákladov.

Príjmy zdravotného systému sú kalkulované na základe štruktúry a počtu prispievateľov do zdravotného systému, stanoveného percentuálneho podielu odvodov a výšky základu pre odvody do zdravotného systému⁴. V tomto module sa oddelene počítajú aj príjmy a výdavky Sociálnej poisťovne, opäť na základe štruktúry poistencov a prispievateľov do Sociálnej poisťovne.

6. Modul ekonomiky

Časť vstupov tohto modulu reprezentujú zadané hodnoty parametrov definujúcich reálny rast HDP, rast produktivity práce, mieru inflácie, rast reálnej mzdy a reálnu úrokovú mieru. V tomto module je opäť možné využiť existujúce prognózy vývoja ekonomiky (vypracované mimo tohto modelu) a použiť tak exogénne nastavenia jednotlivých makroekonomických ukazovateľov.

Tento modul je zameraný aj na prognózovanie príjmov a výdavkov verejného rozpočtu. Na strane príjmov kalkuluje príjmy z priamych a nepriamych daní a iné príjmy (opäť je možné dodať vlastné exogénne prognózy); na strane výdavkov počíta výdavky na zdravotníctvo (v module zdravotného rozpočtu), výdavky na dôchodkové zabezpečenie (stanovené na základe prognózy obyvateľstva v dôchodkovom veku a predpokladoch o výške a raste dôchodkov), výdavky na vzdelanie (zohľadňujúce prognózu obyvateľstva v predproduktívnom veku) a iné výdavky. Model výdavkov na zdravotnú starostlivosť teda poskytuje okrem prognózy zdravotného rozpočtu aj približnú prognózu stavu celkového verejného rozpočtu (samozrejme pri exogénnom zadaní ostatných sáld, napríklad schodkov hospodárenia miest a obcí).

⁴ Prognóza príjmov sa opiera o predpoklady o raste miezd – viď *Modul ekonomiky*.

7. Predpoklady prognózy

Prognóza vývoja hospodárenia verejného zdravotného systému je založená na troch základných predpokladoch a na dátovej základni z roku 2005.

Celkový objem prerozdeľovaných prostriedkov v rámci verejného rozpočtu ako pomer k hrubému domácomu produktu (HDP) bude stabilný v rokoch 2005 až 2050. Tento predpoklad je konzistentný s požiadavkou porovnateľnosti prognóz v rámci projektu FP6 AHEAD. Všeobecne sa pri tom predpokladá, že deficit verejných financií v rokoch prognózy nepresiahne 3 % na HDP. Súčasné trendy reformy zdravotného zabezpečenia a poistenia naznačujú odklon od priebežného verejného financovania. Tieto však neboli zohľadnené v prognóze, predovšetkým preto že nie je istá budúcnosť týchto zmien.

Predpokladáme, že ekonomický rast meraný reálnym rastom HDP bude konvergovať k priemernému tempu rastu eurozóny a že slovenská ekonomika skonverguje k priemeru európskej únie, merané v HDP na obyvateľa, medzi rokmi 2030 až 2040. Tento predpoklad vychádza zo štúdie Performance and Perspective of the European Union as Seen by the New Member States – Post-Accession Monitoring Conference.[7]. Očakávame relatívne vysoký ekonomický rast v nasledujúcom desaťročí, ktorý bude podporovaný prílevom priamych zahraničných investícií, rastom zamestnanosti, politickou stabilitou stredo-európskeho regiónu a komparatívnymi výhodami slovenskej ekonomiky oproti pôvodnej EU12. Po konvergenčnej perióde predpokladáme, že slovenská ekonomika stratí niektoré zo svojich komparatívnych výhod a preto celkový reálny ekonomický rast sa spomalí. Preto rast hospodárstva predpokladáme na úrovni 2,5 % ročného reálneho rastu, rovnako ako je prognózovaný priemer EU.

Posledným východiskovým predpokladom je predpoklad zmeny štruktúry populácie. Táto zmena je následok dramatického poklesu pôrodnosti v 90-tych rokoch 20. storočia a nárastu stredného veku dožitia pri narodení, resp. starnutia populácie. Táto zmena sa najviac prejaví na trhu práce. Pokles v celkovej fertility medzi rokmi 1990 až 2050 je výsledkom transformácie ekonomiky z plánovanej na trhovo orientovanú ekonomiku. Nestabilný ekonomický vývoj spojený so šokovými reformami viedli k poklesu reálnych príjmov domácností a reálneho domáceho produktu. Ako výsledok tohto vývoja celková populácia vo veku 5 až 20 rokov je výrazne nižšia ako v minulosti a toto bude mať dlhodobý vplyv na vývoj ponuky práce na trhu práce. Očakávame, že celkový podiel práceschopného obyvateľstva vo veku 20 až 62 rokov na celkovom obyvateľstve začne medzi rokmi 2015 – 2018 klesať.

Rast reálneho HDP, produktivity práce, participácie na trhu práce, nezamestnanosti a mzdy sú kľúčové premenné, ktoré majú najväčší dopad na vývoj príjmov zdravotného sektora. Starnutie populácie a s tým spojené štruktúrne zmeny v populácii, pôrodnosť, rast miezd zdravotného personálu, miera využitia zdravotných zariadení a celkový stav zdravotného personálu majú najväčší vplyv na vývoj výdavkov zdravotníckeho rozpočtu. Spolu s dôchodkovým systémom tvoria zdravotné výdaje najväčšie položky verejného rozpočtu.

Demografické premenné

Rast populácie najviac závisí od vývoja miery pôrodnosti a úmrtnosti. V rokoch 2005 až 2015 očakávame priemerný medziročný rast populácie na úrovni 0,06 %. Po roku 2015 predpokladáme že miera úmrtnosti sa natrvalo zvýši nad mieru pôrodnosti a preto medzi rokmi 2015 až 2030 očakávame priemerný medziročný pokles celkovej populácie na úrovni 0,2 % a po roku 2030 dokonca na úrovni 0,5 %. Predpokladáme, že vek dožitia pri narodení

u mužov sa zvýši zo 70 rokov v roku 2003 na 73 do roku 2015. Do roku 2030 očakávame že vzrastie na 75 rokov a do roku 2050 na 77 rokov. Rovnako predpokladáme rast veku dožitia pri narodení u žien a to nárast z 78 rokov v roku 2003 na 80 do roku 2015 a do roku 2030 na 82 rokov. Do roku 2040 očakávame mierny nárast na 83 rokov a potom stabilizáciu na tejto úrovni. Očakávaný vývoj veku dožitia v rokoch 2005 až 2050 je odvodený od predpokladu stabilnej zdravotnej starostlivosti a zabezpečenia dostatočnej výživy obyvateľstva. Celková miera plodnosti klesala od osemdesiatych rokov 20. storočia a predpokladáme že tento trend narazí na svoje dno v roku 2010 s mierou 1,19. Základným dôvodom tohto vývoja je celková ekonomická recesia ku koncu 80-tych rokov minulého storočia, a následne veľký prepád v príjmoch obyvateľstva v období prechodu z plánovanej ekonomiky na trhovo orientovanú v rokoch 1989 – 1993. Predpokladáme, že nestabilný ekonomický vývoj, vysoká nezamestnanosť počas reformného obdobia a prechod na trhovú ekonomiku sú kľúčové faktory poklesu miery plodnosti. Po roku 2000 oživenie bankového sektora spolu s daňovou reformou uskutočnenou v roku 2004 viedli k nárastu ekonomickej výkonnosti a k súčasnemu vysokému ekonomickému rastu. Preto očakávame, že rast reálnych príjmov obyvateľstva a vysoký ekonomický rast v rokoch 2005 až 2015, spojený s poklesom nezamestnanosti budú viesť k obratu vývoja miery plodnosti. Preto očakávame, že v rokoch 2010 až 2030 táto miera narastie z 1,19 na 1,43 a následne do roku 2050 na úroveň 1,71. Napriek tomu očakávame že celkový trend vývoja novonarodených jedincov bude záporný z medziročným poklesom -0,008 %. Tento predpokladaný vývoj súvisí predovšetkým s už spomínaným vývojom v predchádzajúcich obdobiach a teda celkovým poklesom populácie žien vo veku 15 až 49 rokov, kvôli nízkej miery plodnosti medzi rokom 1990 až 2005, ktorá výrazne pod úroveň 2,1, teda miery náhrady populácie. Nepredpokladáme, že by sa miera plodnosti dostala nad úroveň 2,1 do roku 2050. V porovnaní s prognózami, ktoré vykonal INFOSTAT [8] sa naša prognóza najviac blíži k strednému variantu. V prognózach INFOSTATU, mieru úhrnnej plodnosti slovenskej populácie očakávajú na úrovni 1,11 v roku 2005 a 1,4 v roku 2050 pre veľmi nízky scenár, 1,19 v roku 2005 a 1,7 v roku 2050 pre stredný scenár a 1,26 v roku 2005 a 2,1 v roku 2050 pre veľmi vysoký scenár. Na porovnanie vývoja očakávaného veku dožitia pri narodení v prognóze INFOSTATU predpokladajú vek dožitia 69,9 v roku 2005 a nárast na 75 rokov v roku 2050 pre veľmi nízky scenár, 70,4 v roku 2005 a 77,1 v roku 2050 v strednom scenári, 70,8 v roku 2005 a 80,3 v roku 2050 vo veľmi vysokom scenári u mužov. U žien predpokladaný vývoj miery dožitia pri narodení prognózuje INFOSTAT na úrovni 77,7 v roku 2005 a 81,7 v roku 2050 vo veľmi nízkom scenári, 78,3 v 2005 a 84,0 v 2050 pre stredný scenár, 78,7 v 2005 a 87,5 v 2050 pre veľmi vysoký scenár. Bližšie detaily našej prognózy sú v tabuľke 1.

T a b u ľ k a 1: Očakávaný demografický vývoj v rokoch 2003 – 2050

	2003	2010	2015	2020	2025	2030	2035	2040	2045	2050
Celkový rast populácie, v %	0,10	0,07	0,01	-0,11	-0,25	-0,35	-0,44	-0,53	-0,61	-0,68
Celková miera plodnosti	1,29	1,19	1,22	1,29	1,36	1,43	1,50	1,57	1,64	1,71
Očakávaný vek dožitia – muži	70	72	73	74	74	75	76	76	77	77
Očakávaný vek dožitia – ženy	78	79	80	80	81	82	82	83	83	83
Pomer (muži : ženy)	0,944	0,949	0,951	0,952	0,950	0,948	0,945	0,941	0,938	0,935

Trh práce

Vývoj trhu práce predpokladáme, že bude najviac ovplyvnený ekonomickým rastom v prvom desaťročí prognózy, následne dôležitým faktorom sa stane aj demografický vývoj. Vysoká miera dlhodobej nezamestnanosti je najväčší problém slovenského pracovného trhu. Predpokladáme, že boj s dlhodobou nezamestnanosťou bude dlhý proces - kvôli nízkej vzdelanosti dlhodobo nezamestnaných, predovšetkým ale kvôli nízkej mobilite nezamestnaných. Je vysoko pravdepodobné, že slovenská ekonomika bude vytvárať dostatok pracovných miest, každopádne vzhľadom regionálne rozloženie, tieto pracovné príležitosti nebudú v regiónoch s najvyššou mierou nezamestnanosti. Predpokladáme, že miera nezamestnanosti klesne na úroveň 6 % do roku 2020 a bude predstavovať takmer všetkých dlhodobo nezamestnaných. Ďalší pokles miery nezamestnanosti nepredpokladáme. Napriek tomu že si uvedomujeme že ekonomika bude v danom období, medzi rokmi 2020 až 2030 vystavená rôznym šokom a cyklom je nesmierne obtiažne prognózovať aký to bude mať vplyv na trh práce. Preto predpokladáme, že prirodzená miera nezamestnanosti sa na Slovensku pohybuje niekde okolo 5 – 6 % a aj prípadné fluktuácie nebudú mať drastický dopad na vývoj miery nezamestnanosti. Preto do konca prognózovaného obdobia predpokladáme priemernú mieru nezamestnanosti na úrovni 6 %.

Medzi rokmi 2005 až 2015 očakávame priemerné tempo rastu zamestnanosti na úrovni 1,35 % a to z dôsledku vysokého rastu ekonomiky a dopytu po práci na trhu. Medzi rokmi 2015 až 2020 očakávame, že celková úroveň ponuky práce bude kulminovať, následne začne pomaly klesať čo bude spôsobené predovšetkým štrukturálnymi zmenami vo vekovej štruktúre obyvateľstva. Tieto zmeny budú mať za následok pokles podielu obyvateľstva vo veku 18 až 62 rokov na celkovej populácii a výrazne začne rásť podiel závislej populácie (predovšetkým dôchodcov) na celkovej populácii.

Predpokladáme že celková participácia na trhu práce narastie zo 69,5 % v roku 2003 na 71,1 % v roku 2015. Rast participácie na trhu práce dosiahne svoj vrchol v roku 2025 na úrovni 73,2 %. Medzi rokmi 2025 až 2050 očakávame mierny pokles participácie na úroveň 71,4 % v roku 2040 a následný opätovný rast na úroveň 72,1 % v roku 2050. Tento vývoj vychádza z predkladanej prognózy obyvateľstva. Predpokladáme, že vo vekovej skupine 15 až 28 rokov dôjde k poklesu participácie na trhu práce v rokoch 2005 až 2050 spojený predovšetkým s cieľom mladých ľudí študovať a dosiahnuť vysokoškolské či univerzitné vzdelanie. Na druhú stranu predpokladáme najvyšší nárast participácie na trhu práce vo vekovej skupine 50 až 65 rokov. Tento vychádza jednak s legislatívnych zmien, posun veku odchodu na dôchodok u žien, ale aj ochota obyvateľstva pracovať dlhšie s cieľom zabezpečiť si vyšší príjem na dôchodku z kapitalizačných dôchodkových pilierov.

Vývoj celkovej miery zamestnanosti je závislý od predpokladaného vývoja participácie na trhu práce a ekonomického rastu. Očakávame celkový rast miery zamestnanosti to z 57,6 % v roku 2003 na 65,2 v roku 2015. Po roku 2015 očakávame spomalenie tohto rastu. Miera zamestnanosti dosiahne vrchol v roku 2025 na úrovni 69,4 %. Po roku 2025 očakávame mierny pokles miery zamestnanosti, ktorá dosiahne spodnú hranicu na úrovni 68,0 % v roku 2040. Následne očakávame mierny rast a do roku 2050 by mala miera zamestnanosti dosiahnuť 68,9 %. Celkový vývoj miery participácie, zamestnanosti a nezamestnanosti zachytáva nasledujúca tabuľka 2

T a b u l' k a 2: Očakávaný vývoj trhu práce, Baseline scenár

	2003 (base year)	2010	2015	2020	2025	2030	2035	2040	2045	2050
Miera participácie – celková, v %	69,5	70,0	71,1	72,4	73,2	73,1	71,9	71,4	71,7	72,1
Miera participácie – muži, v %	76,3	76,4	77,0	77,5	77,5	76,9	75,5	74,9	75,0	75,3
Miera participácie – ženy, v %	57,3	59,3	60,8	62,4	63,9	64,4	63,7	63,3	63,4	63,7
Rast zamestnanosti, v %	0,8	0,8	0,5	-0,4	-0,4	-0,7	-1,1	-1,3	-1,4	-1,2
Miera nezamestnanosti, v %	18,8	11,7	8,9	6,0	6,0	6,0	6,0	6,0	6,0	6,0
Miera zamestnanosti 15-64 – celková, v %	57,6	62,0	65,2	68,6	69,4	69,3	68,3	68,0	68,4	68,9
Miera zamestnanosti 15-64 – muži, v %	63,0	67,8	70,6	73,5	73,6	73,1	71,9	71,6	72,0	72,4
Miera zamestnanosti 15-64 – ženy, v %	52,2	56,2	59,7	63,6	65,2	65,5	64,6	64,4	64,8	65,5

Ekonomické premenné

Očakávaný ekonomický vývoj je v strednodobom horizonte modelovaný ekonometrickým modelom ECM-ISWE06q1, modelom s korekčným členom a to pre roky 2006 až 2013. Prognóza je položená na predpoklade vysokého ekonomického rastu v rokoch 2006 až 2008, s vrcholom v roku 2007. Tento predpoklad je odvodený z vysokého prílivu priamych zahraničných investícií a celkovej tvorby kapitálu ako aj rastu konečnej spotreby domácností, vďaka rastu reálnych príjmov. Rovnako prognózované zníženie deficitu zahraničného obchodu prispieva k vysokému rastu. Priemerný rast HDP počas rokov 2006 až 2013 očakávame na úrovni 6,0 % za rok. V porovnaní s prognózami iných inštitúcií (OECD, IMF, Európska komisia) a aktuálny vývoj ide o optimistické predpoklady⁵. Po tomto období rýchleho rastu predpokladáme jeho spomalenie, pričom v roku 2015 očakávame rast pod úrovňou 5,0 %. Toto spomalenie rastu odvodzujeme s postupnej straty komparatívnych výhod slovenskej ekonomiky a približovaním sa výkonnosťou ekonomiky k priemeru európskej únie. Do roku 2030 očakávame, že priemerné tempo rastu HDP ostane nad úrovňou 3,0 % za rok. Vzhľadom na vysokú neistotu prognózy pre ďalšie roky, predpokladáme že tempo ekonomického rastu s konverguje na úroveň 2,5 % za rok do roku 2050.

Možná expanzia slovenskej ekonomiky v rokoch 2006 až 2015 môže byť predovšetkým podporená nízkymi úrokovými mierami, rastom zamestnanosti, priamymi zahraničnými investíciami a rovnako aj oživením domácich investícií. Liberalizácia pracovného trhu a zníženie daňového zaťaženia v rokoch 2004 a 2005 sú ďalšími faktormi, ktoré podporia vysoký ekonomický rast. Vstup do eurozóny môže tiež pozitívne ovplyvniť vývoj ekonomiky. Predpokladáme, že priemerná inflácia v rokoch 2006 – 2015 bude na úrovni 3,1 %, pričom v rokoch 2007 až 2009 predpokladáme jej pokles na úroveň 2,5 %, vzhľadom na plnenie predvstupových kritérií. Po vstupe do eurozóny bude Slovensko musieť dodržiavať pakt stability a rastu, každopádne napriek tomu očakávame mierne fiškálnej disciplína spojené s mierne vyššou infláciou a rastom deficitu. Rovnako prognózovaný vysoký rast konečnej spotreby domácností bude pôsobiť na rast inflácie počas celého obdobia rokov 2006 až 2015. Medzi rokmi 2015 až 2030 očakávame infláciu na úrovni 3,0 % s miernym trendom klesania a do roku 2050 priemernú infláciu na úrovni 2,4 %.

⁵ Táto práca vychádza z konzistentných predpokladov medzinárodného projektu AHEAD. Tieto predpoklady aktuálna hospodárska kríza nepotvrdzuje, avšak dlhodobé závery sa tým nenarušujú a potvrdzujú sa riziká prognózy.

Rast produktivity práce a miezd sú úzko spojené, preto predpokladáme rovnaké trendy vývoja týchto ukazovateľov. V rokoch 2006 až 2015 očakávame priemerné tempo rastu reálnej produktivity práce na úrovni 4,7 % za rok a reálnych miezd na úrovni 4,2 %. Rast produktivity práce bude podporený vysokými investíciami do pracovnej sily, celoživotným vzdelávaním a rekvalifikáciou nezamestnaných. Po roku 2015 očakávame pokles tempa rastu produktivity práce a rovnako rastu reálnych miezd. Medzi rokmi 2015 až 2030 očakávame priemerné tempo rastu produktivity práce na úrovni 3,5 % a reálnej mzdy 3,2 %. Toto spomalenie je spojené so spomalením rastu ekonomiky ako aj celkovou úrovňou vzdelanosti pracovnej sily, ktorá dosiahne úrovne západných krajín a nebude už profitovať z dovozu manažérskych zručností a know-how zo zahraničia. Po roku 2030 prognózujeme ďalšie spomaľovanie rastu produktivity práce ako aj reálnych miezd to na úroveň 2,3 % resp. 2,2 % v roku 2050, i keď si uvedomuje, že takáto prognóza je spojená s vysokou mierou neistoty. Ekonomická prognóza je detailnejšie zobrazená tabuľka 3.

Tabuľka 3: Očakávaný ekonomický vývoj, baseline scenár, v %

	2003 (base year)	2010	2015	2020	2025	2030	2035	2040	2045	2050
Reálny rast HDP	4,5	6,2	5,1	4,1	3,2	3,1	2,9	2,8	2,6	2,5
Deflátor HDP	5,1	2,9	3,0	2,9	2,8	2,6	2,5	2,3	2,2	2,0
Nominálny rast HDP na obyvateľa	9,0	9,2	8,3	7,2	6,3	6,1	5,9	5,7	5,5	5,3
Rast produktivity práce	3,6	5,3	4,6	3,7	3,0	2,9	2,7	2,6	2,4	2,3
Rast reálnych miezd	-2,1	4,3	4,0	3,4	2,9	2,8	2,6	2,5	2,3	2,2
Inflácie, meraná CPI	8,6	2,7	3,1	3,0	3,0	2,8	2,6	2,4	2,2	2,0
Priemerná reálna úroková miera	-1,0	2,4	2,2	2,7	3,0	2,9	2,8	2,7	2,6	2,5

8. Projekcia príjmov a výdavkov zdravotného systému

Projekcia príjmov a výdavkov zdravotného systému je rozdelená na dve časti. Prvá obsahuje Baseline scenár, základný scenár, ktorý považujeme za najpravdepodobnejší vývoj financovania a hospodárenia zdravotného systému. Tento scenár vychádza z ekonomických prognóz, očakávaní a predpokladov ktoré boli popísané v predchádzajúcej kapitole. Následne sa zaoberáme analýzou senzitivnosti, teda citlivosti hospodárenia zdravotného systému na zmenu demografických či ekonomických predpokladov a očakávaní.

Baseline scenár

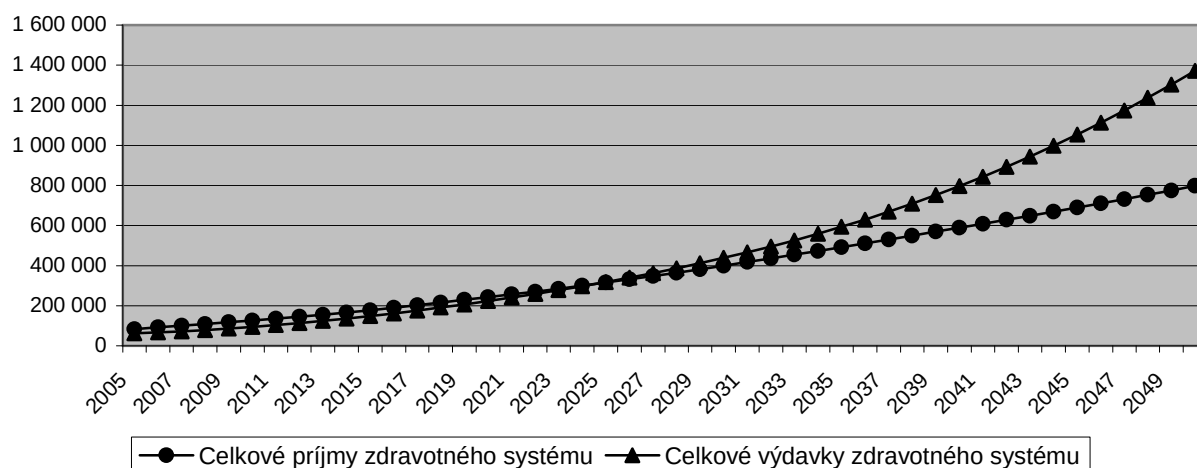
Tento scenár predstavuje najpravdepodobnejší vývoj hospodárenia a financovania zdravotného systému. Prognózovanie vývoja hospodárenia bolo postavené na všetkých dostupných údajov do roku 2005 vrátane. Na základe ekonomickej a demografickej prognózy prognóza vývoja hospodárenia zdravotného systému naznačuje, že v strednodobom horizonte je možné zachovať súčasný model financovania, avšak v dlhodobom horizonte je súčasný model finančne neudržateľný. Preto pre obdobia rokov 2006 až 2025 sa dá očakávať mierna prebytkovosť hospodárenie zdravotného systému avšak po tomto roku systém upadá do dlhohodej narastajúcej deficitnosti. Toto súvisí jednak z demografickým vývojom, ale zároveň aj s predpokladaným vývojom slovenskej ekonomiky. Rovnako odhadnutá miera zaťaženia zdravotného systému je ovplyvnená krátkodobým poklesom pacientov, ktorý spôsobený zavedením platieb za zdravotnú starostlivosť. Autori si uvedomujú tieto

nedostatky, predovšetkým je otázne zrušenie poplatkov za zdravotnú starostlivosť a tiež či tieto poplatky povedú aj k dlhodobému poklesu zaťaženia systému. Napriek týmto nedostatkom však prognóza poukazuje na veľmi pravdepodobný vývoj, a pri zmene orientácie hospodárskej politiky a odklonu od liberalizácie financovania zdravotného systému, je možné očakávať len horší vývoj finančného hospodárenia, na čo poukazuje aj analýza senzitivnosti.

Na základe týchto predpokladov a prognóz ekonomického vývoja prognózujeme rast príjmov zdravotného systému na úrovni 7,8 % za rok a výdavkov 9,2 % medzi rokmi 2005 až 2015. Toto bude mať za následok strednodobú prebytkovosť systému avšak znižujúcu sa. Po tomto období nastane pokles v príjmoch systému súvisiaci s prognózovaným ekonomickým aj demografickým vývojom. Medzi rokmi 2015 až 2035 očakávame priemerný rast príjmov zdravotného systému na úrovni 5,2 % a výdavkov na úrovni 7,2 %. V tomto období sa systém dostane do deficitnosti ktorá bude naďalej narastať. V rokoch 2036 až 2050 očakávame rast príjmov priemernej na úrovni 3,3 % za rok a výdavkov 5,7 %. Vysoký rast výdavkov zdravotného systému spôsobia predpokladané demografické zmeny a starnutie obyvateľstva, pričom najdôležitejší je nárast miery závislosti neproduktívneho obyvateľstva na obyvateľstve v produktívnom veku. Medzi rokmi 2005 až 2015 narastie obyvateľstvo v poproduktívnom veku z 11,9 % v roku 2005 na 13,8 % v roku 2015. Najvyšší nárast však očakávame po roku 2015 do roku 2035 keď tento podiel na celkovej populácii narastie na 21,4 % a do roku 2050 predpokladáme nárast na 27,9 %. Takýto vývoj má dramatický vplyv na výdavkovú stránku systému, ktorá narastá a zároveň aj na príjmovú, ktorá klesá. Pomer produktívneho obyvateľstva na celkovom bude stabilný do rokov 2010 až 2015 následne však očakávame pokles z priemerných 50 % v týchto rokoch na 46,5 % do roku 2035 a 41,7 % do roku 2050. Tento vývoj spojený s očakávaným znížením ekonomického rastu bude mať najväčší dopad na pokles rastu príjmov systému. Celková deficitnosť zdravotného systému predpokladáme, že dosiahne do roku 2050 deficit 2,3 % HDP za rok. Pri nezmenenej filozofii financovania zdravotného systému by bolo nutné zvýšiť zdravotné odvody z hrubej mzdy na úroveň cez 21,5 % zo súčasných 14,0 %.

T a b u l' k a 4: Výsledky prognózy, Baseline scenár

	2010	2015	2020	2025	2030	2035	2040	2045	2050
Celkové zdravotné výdaje ako podiel na HDP, v %	5,0	5,1	5,4	5,6	5,8	6,1	6,4	6,6	6,9
Deficit(-), prebytok (+) ako podiel na HDP, v %	1,4	0,8	0,4	0,0	-0,4	-0,9	-1,4	-1,9	-2,3
Deficit(-), prebytok (+) (v mld.Sk)	31,9	28,7	19,7	-1,9	-38,7	-101,0	-207,1	-364,5	-573,8
Verejné zdravotné výdavky ako podiel na celkových verejných výdavkoch, v %	10,9	11,4	12,0	12,5	13,0	13,5	14,0	14,6	15,1
Celkové verejné výdavky ako podiel na HDP, v %	38,2	37,6	37,4	37,4	37,3	37,3	37,2	37,2	37,2



G r a f 1: Prognózované príjmy a výdavky zdravotného systému

9. Analýza senzitivity

Predkladaná analýza senzitivity predstavuje možné variantné scenáre vývoja financií zdravotného systému v rokoch 2005 až 2050. Analýza bola vykonaná z ohľadom na možnú zmenu vývoja demografických ako aj ekonomických premenných. Z demografických sa jedná o zmenu vývoja očakávaného veku dožitia u mužov aj žien. Predpokladáme, že očakávaný vek dožitia pri narodení môže rásť rýchlejšie, resp pomalšie a teda populácia bude žiť dlhšie alebo kratšie. Tento očakávaný vývoj zachytávajú nasledujúcej tabuľky 5 a 6.

T a b u ľ k a 5: Očakávaný vývoj veku dožitia, muži

	2003 (base year)	2010	2015	2020	2025	2030	2035	2040
Očakávaný vek dožitia, rýchly scenár	70	72	73	74	75	76	77	77
Očakávaný vek dožitia, baseline scenár	70	72	73	74	74	75	76	76
Očakávaný vek dožitia, pomalý scenár	70	71	72	73	73	74	74	75

T a b u ľ k a 6 : Očakávaný vývoj veku dožitia, ženy

	2003 (base year)	2010	2015	2020	2025	2030	2035	2040
Očakávaný vek dožitia, rýchly scenár	78	79	80	81	82	83	83	84
Očakávaný vek dožitia, baseline scenár	78	79	80	80	81	82	82	83
Očakávaný vek dožitia, pomalý scenár	78	79	79	80	80	80	81	81

Celkovo predpokladáme, že rovnako u mužov tak aj u žien by predpokladaný vek dožitia narástol, resp poklesol o jeden až dva roky. Dopad takéhoto vývoja na financovanie

zdravotného systému nie je veľký. Rozdiel vo výdavkoch systému zachytáva nasledujúca tabuľka 7.

T a b u ľ k a 7: Projekcia výdavkov zdravotného systému vzhľadom na rozdielny očakávaný vek dožitia

	2010	2015	2020	2025	2030	2035	2040	2045	2050
Celkové výdavky zdravotného systému – rýchly scenár	93,8	148,6	223,7	319,0	441,3	598,2	804,1	1 066,8	1 392,8
Celkové výdavky zdravotného systému – baseline scenár	93,8	148,4	223,1	317,6	438,4	593,2	796,1	1 053,8	1 371,7
Celkové výdavky zdravotného systému – pomalý scenár	93,7	148,2	222,4	316,4	436,3	589,2	789,1	1 043,1	1 356,9

Celkovo sa dá očakávať, že výdavky zdravotného systému budú len mierne ovplyvnené. Celkové výdavky by vzrástli 1,53 % k roku 2050 pre rýchly scenár, resp. poklesli o 1,07 % pre pomalý scenár v roku 2050. Toto sú len nepatrné rozdiely, napriek tomu je zrejmé že nárast očakávaného veku dožitia a teda celkovo rast veku dožitia má negatívny vplyv na vývoj financovania zdravotného systému.

Ďalší scenár predstavuje zmenu predpokladu vývoja nákladov zdravotného systému. Pre baseline scenár náklady zdravotného systému, teda odmeny zdravotného personálu a ostatné náklady rastú rovnakým trendom ako sa vyvíja tempo rastu nominálneho HDP na obyvateľa. V predkladaných scenároch senzitivnosti, je pre Scenár A, oddelený vývoj nákladových položiek, teda mzdové sa vyvíjajú rozdielne ako ostatné a pre scenár B je nárast nákladov odvodený od vývoja nominálnej produktivity práce. Výsledky týchto scenárov sú znázornené v nasledujúcej tabuľke 8.

T a b u ľ k a 8: Prognóza výdavkov systému vzhľadom na rozdielne predpoklady tvorby nákladov

	2010	2015	2020	2025	2030	2035	2040	2045	2050
Celkové výdavky zdravotného systému – Scenár A	80,9	111,2	154,1	209,6	279,3	365,0	473,1	604,1	757,8
Celkové výdavky zdravotného systému – baseline scenár	93,8	148,4	223,1	317,6	438,4	593,2	796,1	1 053,8	1 371,7
Celkové výdavky zdravotného systému – Scenár B	86,4	132,9	196,5	282,8	396,7	551,2	767,7	1 058,1	1 421,2

Je zrejmé, že ak by sa všetky zložky nákladov v zdravotnom systéme vyvíjali rovnako tak celkové výdavky by sledovali podobnú trajektóriu. Preto rozdiely medzi baseline scenárom a scenárom B nie sú veľké aj keď vývoj výdavkov podľa tempa rastu produktivity práce by bol v strednodobom horizonte nižší následne by však akceleroval a do roku 2050 by dosiahol vyššiu hladinu. Scenár A predstavuje rast výdavkov kde sú mzdy zdravotného personálu oddelené a rastú rovnako v baseline scenári, ale ostatné výdavky rastú pomalšie. Jedná predovšetkým výdavky na lieky, zdravotný materiál, investície a iné a ich rast dosahuje tretinu rastu mzdových nákladov. Takto získaná prognóza dosahuje signifikantný rozdiel, treba si však uvedomiť, že je málo pravdepodobný takýto vývoj.

Nasledujúce scenáre predstavujú dopad rozdielneho vývoja ekonomických ukazovateľov. Zamerali sme sa na kvantifikáciu dopadu príjmov aj výdavkov vzhľadom na rozdielny očakávaný vývoj miery zamestnanosti, reálnej mzdy a ich kombinácii.

Vytvorené scenáre sú s očakávanou vyššou mzdou (HW scenár), s očakávanou nižšou mzdou (LW scenár), očakávanou vyššou mierou zamestnanosti (HER scenár), nižšou mierou zamestnanosti (LER scenár) ich kombinácie vysoká mzda a rovnako zamestnanosť (HWHHER scenár), vysoká mzda nízka zamestnanosť (HWLER), nízka mzda vysoká zamestnanosť (LWHER), a nízka mzda nízka zamestnanosť (LWLER).

Predpoklad a očakávanie rozdielneho vývoja mzdy je oprávnený. Celkovo však baseline scenár zachytáva najpravdepodobnejší vývoj. Nato aby došlo k rozdielnemu vývoju mali by byť naplnené niektoré predpoklady a iné nie. Na to aby mzda rástla rýchlejšie ako predpokladáme muselo by dôjsť k podstatnému nárast produktivity práce. To je možné za predpokladu, že dôjde v sledovanom období k podstatným štrukturálnym zmenám v ekonomike. Predovšetkým by musela ekonomika pritiahnuť alebo vytvoriť technologicky vysoko náročné odvetvia produkcie a odvetvia v ktorých výrazne rastie pridaná hodnota. Toto by bolo možné za predpokladu, že celkové výdavky na výskum a vývoj narastú, jednak v súkromnom sektore ale aj vládne výdavky. Pre takýto vývoj je nesmierne dôležité aby ekonomické prostredie EU ostalo dlhodobo stabilné.

Naproti to scenár s očakávaným nižším rastom priemernej mzdy vychádza z opačných predpokladov. Jednak ekonomické prostredie európskej únie nebude stabilné a naviac slovenská ekonomika nebude schopná vytvoriť pracovné miesta s vysokou pridanou hodnotou a produktivitou práce. To by znamenalo orientáciu ekonomiky na prácu náročnú produkciu y nízkou pridanou hodnotou. Taktiež by to znamenalo, že podiel služieb na celkovej produkcii nedosiahne priemeru vyspelých západných krajín. Celkovo predpokladáme odklon od základného scenára vo vývoji priemernej mzdy na úrovni 1,5 p.b. v oboch scenároch.

Zvyšovanie či znižovanie miery zamestnanosti oproti baseline scenáru je odvodené od predpokladu nárastu či poklesu ochoty pracovať. Celková demografická prognóza je totožná s prognózou v baseline scenári.

Na to aby rástla ochota pracovať a teda sa zvýšila miera zamestnanosti je nutné aby v ekonomike rástli reálne mzdy, na druhú stranu aby miera uspokojenia s príjmom nedosiahla bod, keď rodine postačí jeden príjem. Predpokladáme, že práve štrukturálne zmeny v ekonomike a posun k výrobe náročnej na vysoko vzdelanú pracovnú silu spolu s konvergenciou podielu služieb na celkovej produkcii k priemeru EU, môže viesť k ochote viac pracovať. Ďalším dôležitým aspektom je predlžovanie veku dožitia a rovnako aj odchodu do dôchodku. Preto práve vo vekovej skupine 55 – 64 rokov predpokladáme najvyšší nárast participácie na trhu práce to zo súčasných 27 % na 62 % v roku 2050. Ďalšou vekovou skupinou kde predpokladáme výrazný rast je 45 – 54 ročný, kde predpokladáme nárast zo súčasných 82 % na 97 %. Jedinou vekovou skupinou s poklesom participácie sú 15 – 24 ročný kde predpokladáme uprednostnenie vzdelávania pred rýchlym nástupom do pracovného pomeru. Celkovo v scenári s vyššou mierou zamestnanosti predpokladáme nárast participácie o dodatočných 9 p.b. za všetky vekové skupiny.

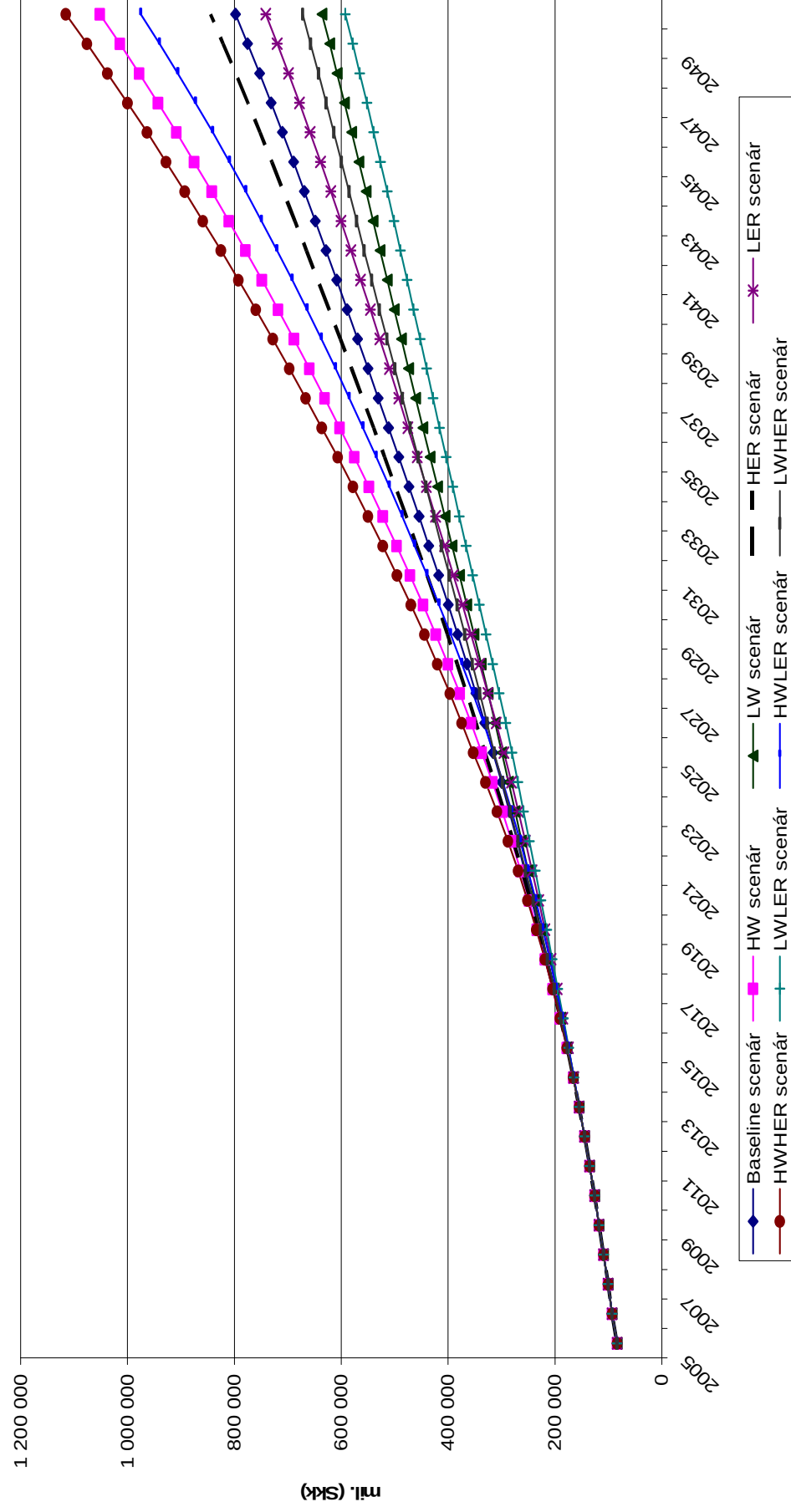
Scenár z nízkou mierou zamestnanosti vychádza z predpokladu, že celková participácia na trhu práce ostane dlhodobo takmer nezmenená. Oproti základnému scenáru, baseline, predpokladáme, že participácia poklesne predovšetkým ku koncu sledovaného obdobia. Celkovo však oproti dnešnému stavu participácia vzrastie o 2 p.b. do roku 2050. Predpokladaný vývoj participácie bude kulminovať medzi rokmi 2015 – 2025, potom začne stagnovať a ani nízka nezamestnanosť a predpokladaný vysoký dopyt po práci nebudú výrazne ovplyvňovať participáciu. Najvyšší nárast očakávame vo vekovej skupine 55 – 64 rokov a to na úroveň 47 % zo súčasných 27 %.

Kombinácie predpokladov vývoja miezd a miery zamestnanosti viedli k tvorbe štyroch scenárov. V týchto scenároch sa prejavujú synergické efekty predpokladov a dva z nich predstavujú hornú a spodnú hranicu očakávaného vývoja. Jedná sa o scenár s vysokými mzdami a zároveň zamestnanosťou (HWHER), horná hranica a scenár s nízkymi mzdami a rovnako nízkou mierou zamestnanosti tvorí dolnú hranicu (LWLER).

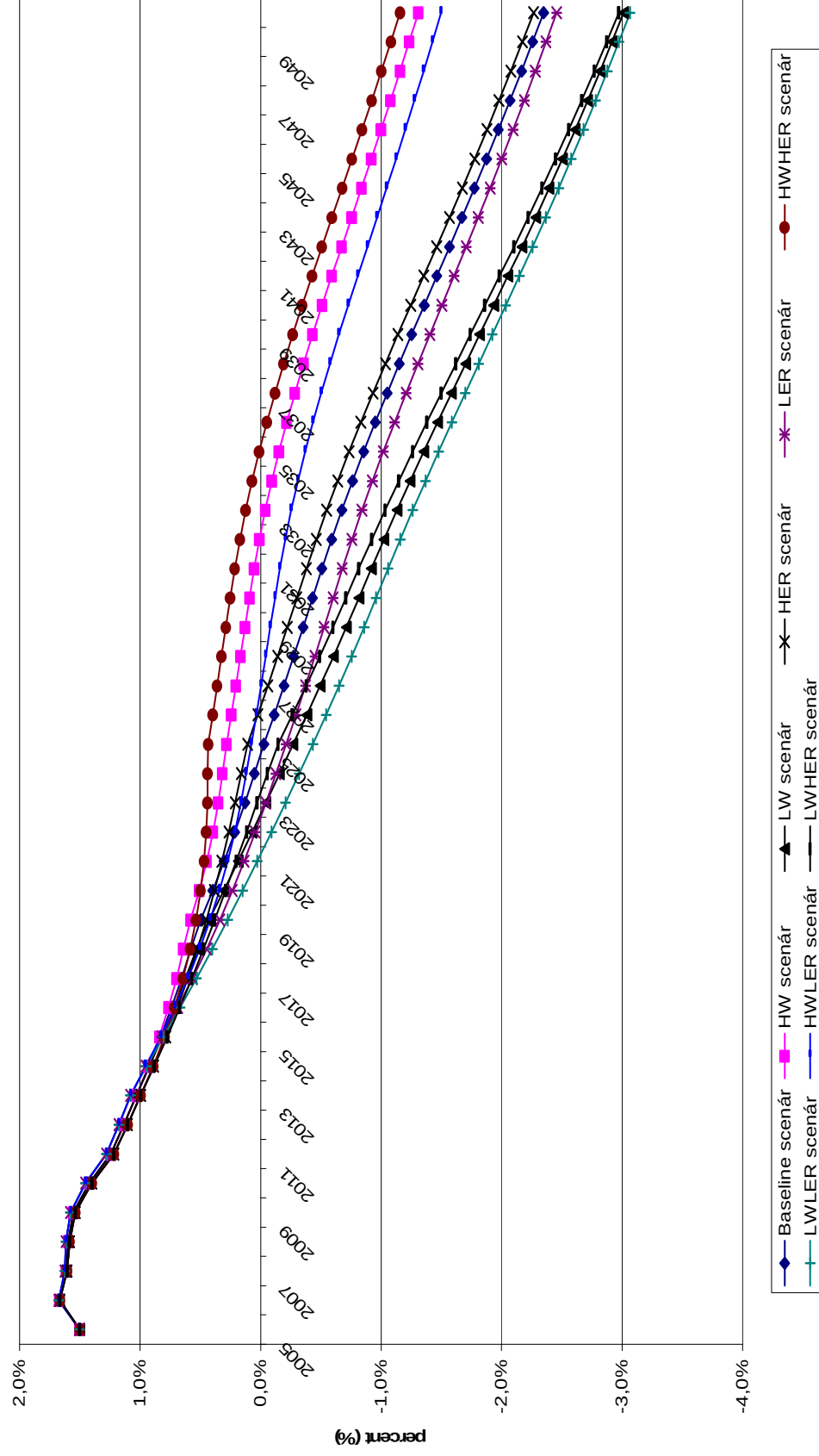
Ekonomický vývoj, ktorý naznačujú tieto scenáre je značne nepravdepodobný a teda aj vývoj hospodárenia zdravotného sektora. Horná hranica, teda scenár HWHER predstavuje synergiu kladných očakávaní a zároveň by ekonomika nemala byť vystavená šokom či dlhodobým negatívnym efektom. Naproti tomu scenár LWLER, predstavuje opak. Takýto ekonomický vývoj sa dá očakávať len pri veľmi neuváženej hospodárskej politike, spojenej navyše s nepriaznivým ekonomickým vývojom EÚ a zahraničia.

Scenáre HWLER a LWHER predstavujú scenáre s naplneniami niektorých očakávaní avšak nie všetkými. Pre scenár HWLER predpokladáme, že očakávania ohľadne vývoja reálnej mzdy budú naplnené, ale toto bude viesť k nižšej ochote pracovať, predovšetkým kvôli uspokojivej úrovni príjmu domácnosti od jedného pracujúceho. Takýto vývoj sa dá očakávať ak sociálna situácia nebude vnímaná negatívne a zároveň ekonomický rast a rast produktivity práce budú tvoriť dostatočné zdroje na rast miezd.

Scenár LWHER predstavuje obdobu predchádzajúceho, s opačnými predpokladmi. Predpokladaný vývoj rastu reálnych nebude naplnený, naopak ale bude viesť k rastu ochoty pracovať. Nenaplnenie očakávaní rastu reálnej mzdy a teda jej nižšie tempo rastu, spojené zároveň s limitovanými prímami verejného rozpočtu, ktoré nebudú postačujúce na tvorbu aktívnej a ani pasívnej sociálnej politiky, môžu viesť k zvýšeniu ochoty pracovať. Predpokladáme, že ekonomika bude vytvárať dostatočný dopyt po práci a teda jednou z možností ako si zlepšiť životné podmienky v takejto situácii bude získať dodatočný príjem z práce. Výsledky týchto scenárov sú graficky znázornené na nasledujúcich grafoch (Graf 2 a 3).



G r a f 2: Príjmy zdravotného systému, variantné scenáre



G r a f 3: Prebytok (+), Deficit (-) zdravotného systému ako podiel na HDP

1. Záver

V rámci vedecko-výskumného projektu AHEAD celkom výskumné tímy 18 riešiteľských pracovísk zo 16 krajín v priebehu skoro 4 ročného výskumu v deviatich pracovných balíkoch skúmali starnutie, zdravotný stav a determinanty výdavkov na zdravie v rozširujúcej sa Európskej únii. Mnohé z týchto výsledkov sú relevantné pre Slovensko, avšak predovšetkým časti venované vybratým novým členským krajinám sú pre nás mimoriadne aktuálne.

Pre vývoj scenárov zdravotných výdavkov boli za vybraté nové členské štáty EÚ (Poľsko, Slovensko, Maďarsko, Bulharsko a Estónsko) zostrojené modely, ktoré mali tieto spoločné špecifické črty:

zahŕňali externé faktory zdravotného systému: demografia, pracovný trh, hospodársky vývoj,

- podrobný model zdravotného systému obsahoval príjmovú aj výdavkovú stranu,
- sústredenie na rovnováhu poisťného systému a finančnú udržateľnosť zdravotného systému,
- založené na národných prognózach vývoja demografie, pracovného trhu a makroekonomiky,
- založené na národných systémoch a zvyklostiach používania zdravotníckych služieb,
- výsledky smerujúce na politické odporúčania.

Z výsledkov prognóz jednotlivých sledovaných štátov vyplynuli tieto všeobecné tendencie:

- systematický rast deficitu zdravotných systémov,
- rast deficitov spôsobený tendenciami na oboch stranách: výdavkovej aj príjmovej, veľmi limitovaný rast príjmov, veľmi vysoký rast výdavkov a ako výsledok – systematický rast deficitu,
- udržateľnosť financovania zdravotného systému priamo závisí od ekonomického vývoja - s ukazovateľmi: hospodársky rast, miera zamestnanosti a mzdy,
- výdavková strana priamo závisí od využívania služieb zdravotného systému a ich ceny majú byť podmienené efektívnym fungovaním zdravotného systému.

Zo strednodobého hľadiska je súčasťou model financovania zdravotného systému udržateľný, predovšetkým vďaka vysokému hospodárskemu rastu. Akonáhle však narazí na dve základné bariéry a to pokles hospodárskeho rastu a rast podielu závislej populácie na produktívnej, systém sa rýchlo dostane do deficitnosti. Spolu s očakávaným dlhodobým nepriaznivým populačným vývojom nebude možné zabezpečiť dlhodobú udržateľnosť financovania zdravotníctva. Preto je nutné aby došlo k zmenám priamo v systéme financovania zdravotnej starostlivosti. Ak by nedošlo ku zmene filozofie financovania zdravotného systému bude nutné na udržanie vyrovnaného hospodárenia zvyšovať odvodové zaťaženie pracujúcich, resp rozšíriť odvodovú základňu. Na základe nášho výskumu je zrejmé, že zdravotné odvody bude nutné postupne zvyšovať na úroveň 21 % z hrubej mzdy do roku 2050, aby systém dosiahol vyrovnané hospodárenie. Alternatívne scenáre poukázali na fakt, že jedno percentný dodatočný rast / pokles príjmov vedie k dodatočnému nárastu / poklesu príjmov zdravotného systému na úrovni 0,6 – 0,7 %. Naproti tomu dodatočná jedno percentná zmena miery zamestnanosti vedie k nárastu / poklesu príjmov zdravotného systému o 0,05%. V prípade najpesimistickejšieho vývoja je možné očakávať deficit na úrovni 3 % z HDP, avšak očakávame, že deficitnosť dosiahne najpravdepodobnejšie úroveň deficitu 1,2 – 1,5 % z HDP v roku 2050. Na to by sa zdravotný systém vyhol dlhodobej deficitnosti je nutné aby sa tvorcovia hospodárskej politiky pridržiavali niektorých odporúčaní:

základom pre zvýšenie príjmov zdravotného systému je zvýšenie dynamiky hospodárskeho rastu so stratégiou zamestnanosti a rastu miezd (v súlade s rastom produktivity práce),

- zlepšenie výdavkovej časti závisí od zvýšenia efektívnosti zdravotného systému, administratívnej efektívnosti a výkonných kapacít,
- obsažná zdravotná politika – všetky sociálne poisťné systémy, príjmové a výdavkové časti, verejné zdravie, aktívne – zdravé starnutie a liečebná zdravotná starostlivosť.

2.Literatúra

- [1] KVETAN, V, PÁLENÍK, V, MLÝNEK, M, RADVANSKÝ, M. Starnutie, zdravotný stav a determinanty výdavkov na zdravie v podmienkach Slovenska. Ekonomický ústav SAV VS 1, Bratislava 2007, ISBN 978-80-7144-160-1
- [2] AHEAD Studies Published in the ENEPRI Research Reports Series (available for free downloading from www.enepri.org or <http://shop.ceps.be>)
- [3] No. 31 Health Status and Healthcare Systems in Central and East European Countries: Bulgaria, Estonia, Poland, Slovakia and Hungary, December 2006
- [4] KVETAN, V, PÁLENÍK,V, MLÝNEK, M, RADVANSKÝ, M: Health expenditure scenarios in the new member states country report on Slovakia. ENEPRI research report No. 48, CEPS, Brussels 2007, 27 p. ISBN 978-92-9079-765-4.
- [5] BORS, L. KVETAN, V. PÁLENÍK, V. VOKUN, J.: Construction and Verification of Macroeconomic Model ISWE97q3. Ekon. čas., 46, 1998, è. 3, s. 428-466
- [6] ĎURAŠ, J. KVETAN, V. MLYNEK, M. ONDKO, P. PÁLENÍK, V. RADVANSKÝ, M.: Makroekonomická prognóza ekonomiky SR na roky 2005 – 2009 In. Zborník konferencie: Pohľady na ekonomiku Slovenska 2005, SDSS, Bratislava, 2005
- [7] Performance and Perspective of the European Union as Seen by the New Member States, Šikula M., Kadeřábkoá A., Klvačová E., Mrkaić M., Rezdir R., Okáli I., Vida K., Zeman K., Michalski R., Gaidys V., Sereghyová J., Bakosová K., Kvetan V., Kawecka-Wyrzykowska E., Institute of Economic Research of Slovak Academy of Sciences, 2006, ISBN 80-7144-151-1.
- [8] VAŇO, B.: Population projection of Slovakia until 2050, INFOSTAT – Institute of Informatics and Statistics, Demographic Research Center, Bratislava 2001.

Adresa autorov:

Viliam Páleník, h.doc. RNDr., PhD.

Ekonomický ústav SAV

Šancová 56

811 01 Bratislava

viliam.palenik@savba.sk

Marek Radvanský, Ing.

Ekonomický ústav SAV

Šancová 56

811 01 Bratislava

marek.radvansky@savba.sk

Vladimír Kvetan, Ing.

Ekonomický ústav SAV

Šancová 56

811 01 Bratislava

vladimir.kvetan@gmail.com

Martin Mlýnek, Ing.

Ekonomický ústav SAV

Šancová 56

811 01 Bratislava

martin.mlynek@gmail.com

Analýza determinantov samo hodnotenia zdravotného stavu a využitia zdravotnej starostlivosti

Analysis of the determination of self-assessed health and healthcare utilisation

Vladimír Kvetan, Viliam Páleník

Abstrakt: V práci je analyzovaný súvis medzi samo hodnotením zdravotného stavu populácie Slovenska a využívaním zdravotnej starostlivosti. Pozornosť je venovaná využívaniu zdravotnej starostlivosti, skúmaniu zdravotného stavu na základe samo hodnotenia, vplyvu pohlavia a ďalších faktorov na využitie zdravotných služieb. Využitie pri tom boli verejne dostupné štatistické údaje a prieskumy a následne štatistické testy preukaznosti model typu logit.

Key words: health system, health evaluation, health condition, determinants of health

Kľúčové slová: zdravotný systém, hodnotenie zdravia, zdravotné podmienky, determinanty zdravia

1 Úvod

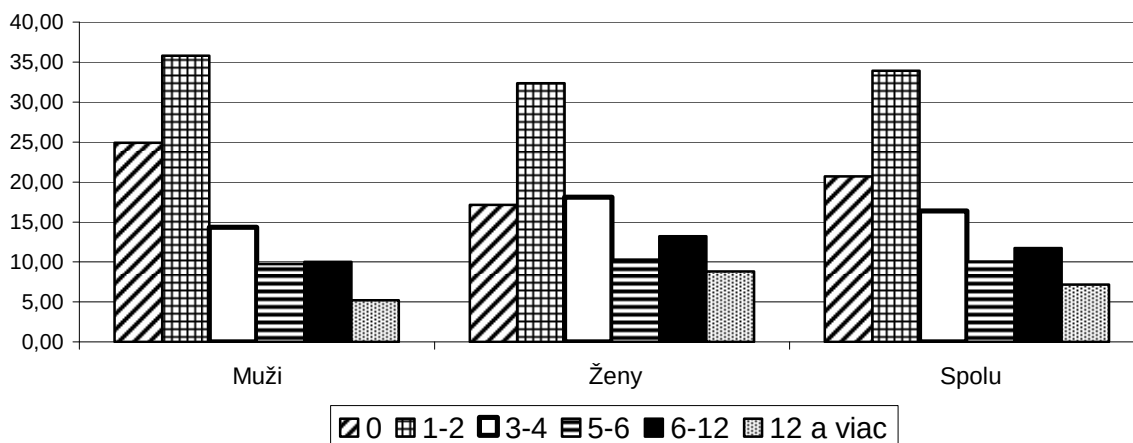
Táto práca vychádza z projektu 6FP AHEAD [10-13], ktorého výsledky relevantné pre Slovensko sú uvedené v [9]. Analýza samohodnotenia zdravotného stavu a využitia zdravotnej starostlivosti bola založená na dátach z monitoringu CINDY. Analyzované boli odpovede na otázku 1 „Koľkokrát ste za ostatných 12 mesiacov navštívili lekára“ a otázku 7 „Ako hodnotíte svoj súčasný zdravotný stav“. Tieto odpovede boli skúmané vzhľadom na demografické a socioekonomické charakteristiky respondenta (vek, pohlavie, ekonomická aktivita a pod.). Pre analýzu samohodnotenia zdravotného stavu boli využité dáta z publikácie Zdravotné uvedomenie a správanie obyvateľov Slovenska.

2 Analýza využitia zdravotnej starostlivosti

Vzhľadom na rozdielnosť vo frekvenciách návštevy u lekára s výnimkou dentistu a pravidelných kontrolách (nie z dôvodu komplikácie) v tehotenstve boli respondenti rozdelení do šiestich skupín: žiadna návšteva, v priemere raz za pol roka, v priemere raz za štvrt' roka, každý druhý mesiac, raz mesačne viac ako raz mesačne. Analýzou týchto skupín sme následne dokázali identifikovať spotrebiteľov s vyšším alebo nižším dopytom.

Pri skúmaní využívania zdravotnej starostlivosti vzhľadom na pohlavie majú muži aj ženy rovnakú frekvenciu, priemerne päť návštev u lekára. Po vylúčení jedného outlayera (muž s 240 návštevami v roku) sa priemerná návšteva zmenila na 4 návštevy u mužov. Je teda zrejmé, že ženy navštevujú lekára priemerne o 1 krát ročne viac ako muži.¹ Ako je možné vidieť z grafu 1 muži majú výrazne odlišnú frekvenciu návštev. Výrazne prevládajú muži so žiadnou alebo jednou či dvoma návštevami u lekára, naopak u žien prevládajú častejšie návštevy 6 – 12 krát ročne a viac ako 12 krát ročne.

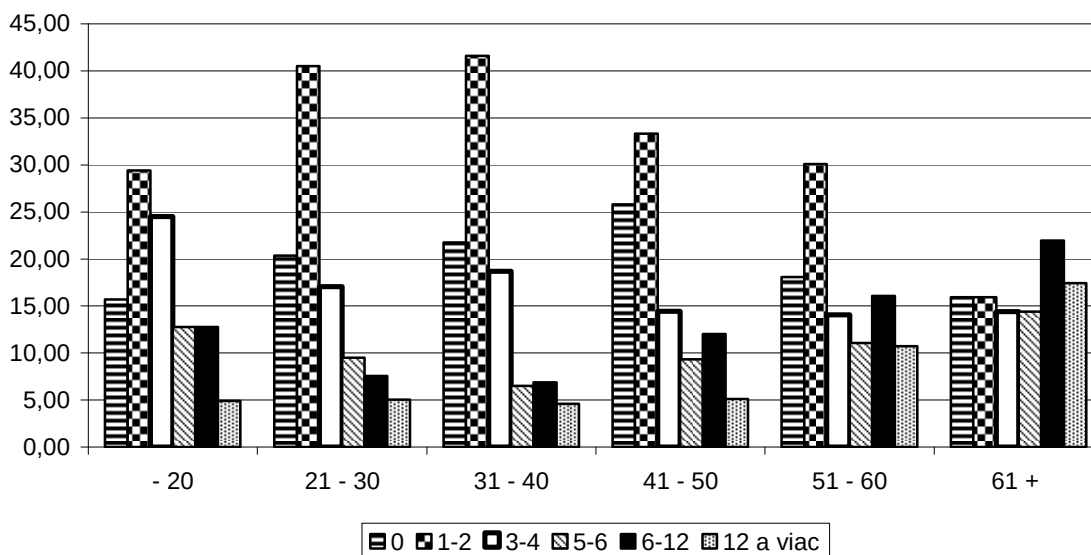
¹ Bol vykonaný T-test, ktorý nepotvrdil hypotézu o rovnosti stredných hodnôt.



Zdroj: CINDY.

G r a f 1 Využitie zdravotnej starostlivosti obyvateľmi Slovenska

Nečakané výsledky poskytla analýza na základe vekových skupín (Graf 1.a). Predpokladali sme postupný nárast návštev u lekára vzhľadom k rastúcemu veku. Stav získaný na základe odpovedí respondentov však odhalil, že najvyšší počet ľudí nenavštevujúcich lekára je vo vekovej skupine 41 – 50 rokov a maximálne dve návštevy u lekára v skupine 31 – 40 rokov. Naopak obyvatelia vo vekovej skupine 20 a menej rokov je priemerná návštevnosť u lekára druhou najvyššou. Frekvencia každý druhý mesiac je veľmi blízka vekovej skupine 51 – 60 a frekvencia každý mesiac je blízka vekovej skupine 41 – 50. Je možné sa domnievať, že návšteva u lekára súvisí s ekonomickou aktivitou a fenoménom „mať čas navštíviť lekára“.

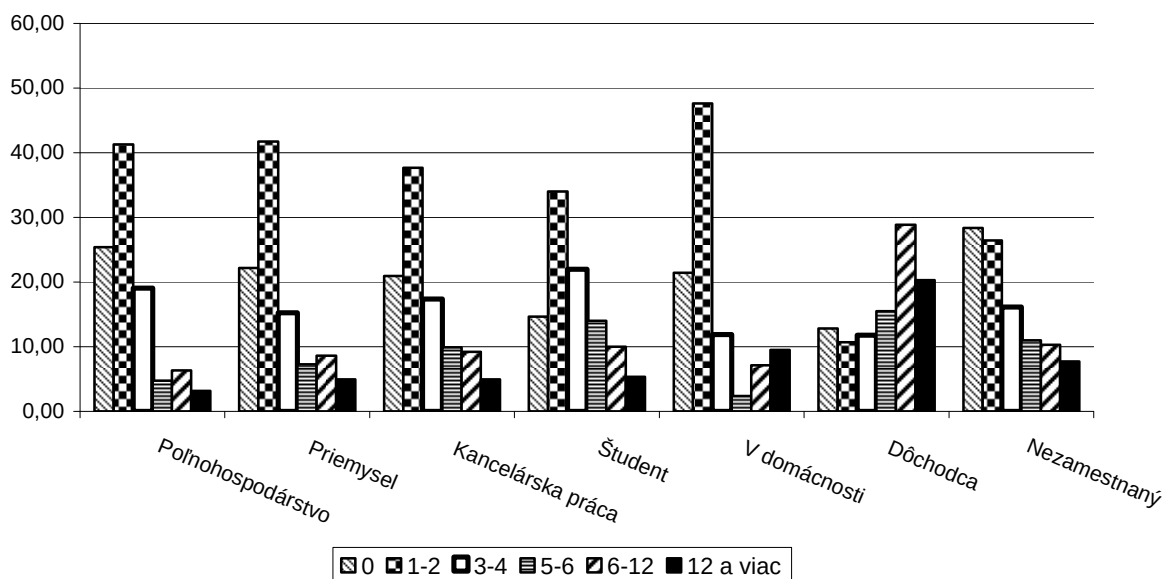


Zdroj: CINDY.

G r a f 1a: Využitie zdravotnej starostlivosti podľa vekových skupín

V realite to znamená, že pracujúci často krát „pretrpí“ menej vážne ochorenie (prechladnutie, bolesť hlavy), aby neabsentoval z práce napríklad z dôvodu výpadku príjmu. Naopak študenti radi využijú akúkoľvek možnosť na únik zo školy. Takisto je však možné dať tento počet návštev do súvisu s možnými objektívnymi zdravotnými komplikáciami pri procese dospievania.

Vyššie zmienené úvahy sú overiteľné pri skúmaní využitia zdravotných služieb na základe skupín ekonomickej aktivity (Graf 2). V skupine študentov je najvyšší pomer jedincov navštevujúcich lekára raz za štvrt'rok. Skupina raz za pol roka je rovnako veľká ako pri nezamestnaných. Najvyšší podiel návštev s najvyššou frekvenciou je pochopiteľne u dôchodcov.

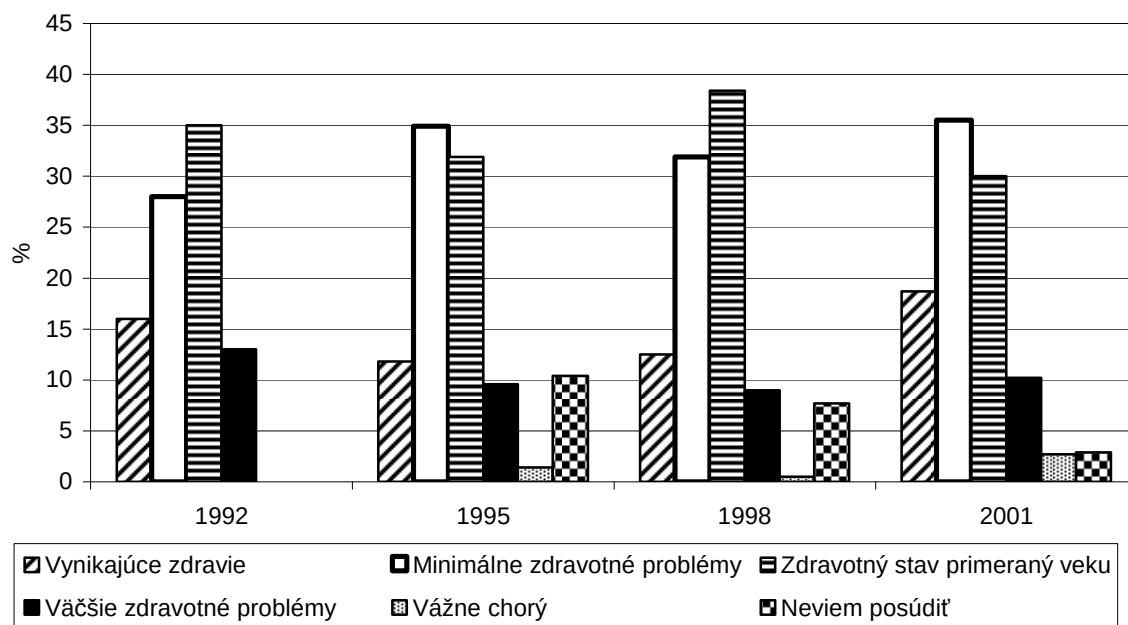


Zdroj: CINDY.

G r a f 2: Využitie zdravotnej starostlivosti podľa skupín ekonomickej aktivity

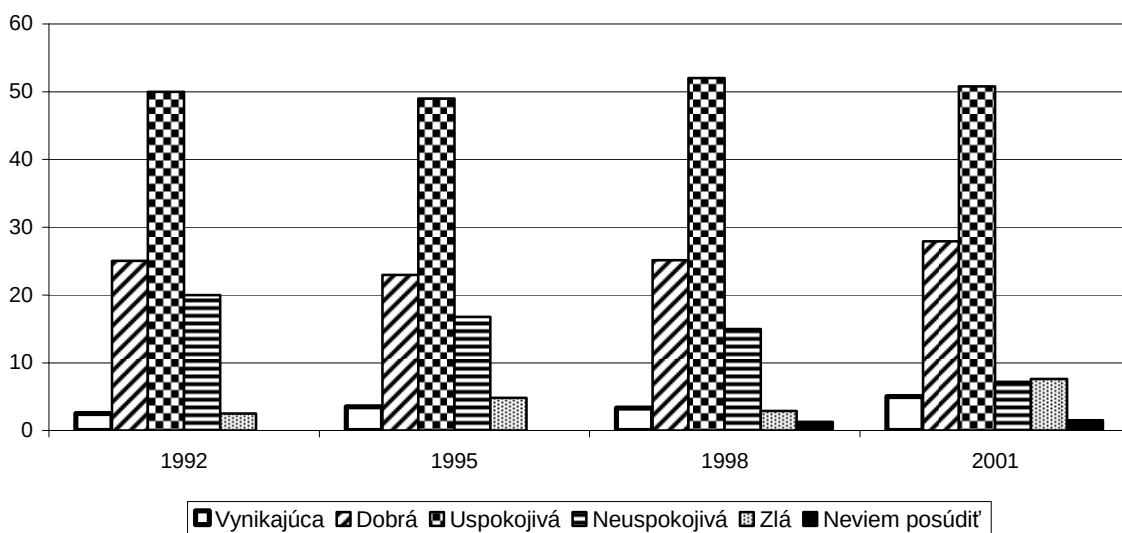
3 Skúmanie zdravotného stavu založené na samohodnotení

Samohodnotenie zdravotného stavu je indikátor, ktorý kvalifikovane popisuje zdravotný stav a subjektívny pocit zdravia respondenta. Štúdia Zdravotné uvedomenie a správanie populácie Slovenskej Republiky [1] je zdrojom dát popisujúcich dynamiku vývoja vnímania individuálneho pocitu zdravia v rokoch 1992 – 2001. Graf 3 poukazuje na postupné vylepšovanie vnímania zdravotného stavu. Výrazne sa zvýšilo percento obyvateľstva prezentujúci svoj zdravotný stav za vynikajúci. Zmeny tiež nastali pri hodnotení „Minimálne zdravotné problémy“ a „Zdravotný stav primeraný veku“. Toto zlepšenie môže byť spojené aj so zlepšením vnímania životnej úrovne (Graf 4). Tu je badateľný posun k vyšším podielom respondentov hodnotiacich svoju životnú úroveň za vynikajúcu alebo dobrú, pričom podiel s odpoveďou „uspokojivá“ ostáva takmer konštantný. Sklon k zovšeobecňovaniu hrá dôležitú úlohu pri individuálnom hodnotení pocitu zdravia, či životnej úrovne. Je predpoklad, že ak sa respondent nachádza v nepriaznivej životnej pozícii a má nízku životnú úroveň (nezamestnanosť, nízka mzda a pod.), tak má vyšší sklon ohodnotiť svoj zdravotný stav horšie. Dôkaz poskytla korelačná analýza. Korelačný koeficient pozitívne hodnoteného zdravotného stavu (vynikajúci, minimálne zdravotné problémy, zdravotné problémy primerané k veku) a pozitívneho hodnotenia životnej úrovne (vynikajúca, dobrá, uspokojivá) bol približne 0,96. Korelácia dvoch najlepšie hodnotiacich skupín zdravia a životnej úrovne bola 0,87.



Zdroj: Zdravotné uvedomenie a správanie populácie Slovenskej Republiky.

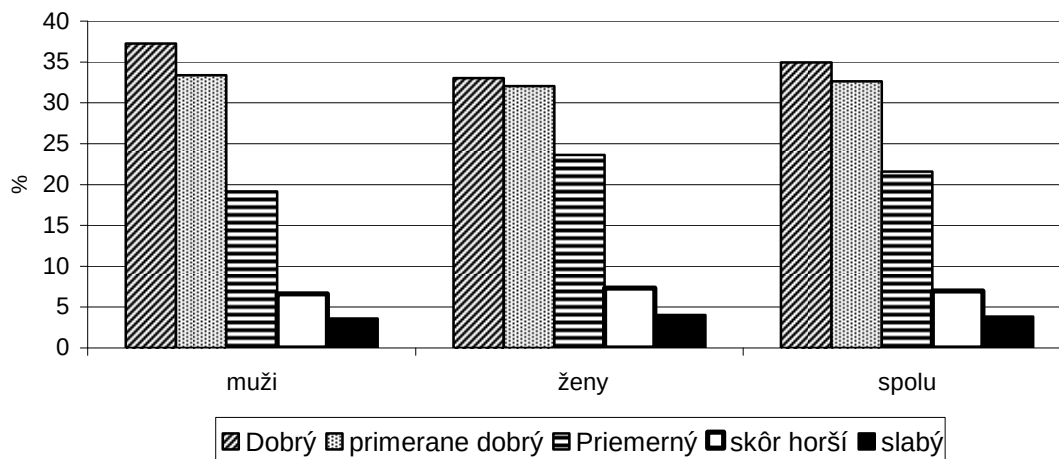
G r a f 3 Hodnotenie zdravotného stavu v SR



Zdroj: Zdravotné uvedomenie a správanie populácie Slovenskej Republiky.

G r a f 4 Hodnotenie životnej úrovne

Analýza založená na dotazníku CINDY poskytuje dáta o samohodnotení zdravotného stavu na základe jednotlivých skupín populácie. Rozdelenia vnímania zdravotného stavu pri oboch pohlaviach sú zošikmené doľava (Graf 5). Najvyššie podiely dosahovali hodnotenia dobrého zdravotného stavu. Muži viac hodnotili svoj stav ako dobrý, na rozdiel od žien, ktoré ho viac hodnotili ako priemerný. Je možné predpokladať istý sklon k opatrnejšiemu individuálnemu hodnoteniu zo strany žien. Sklon je závislý od samotného charakteru prieskumu, ktorý sa sústreďoval na populáciu v ekonomicky aktívnom veku.

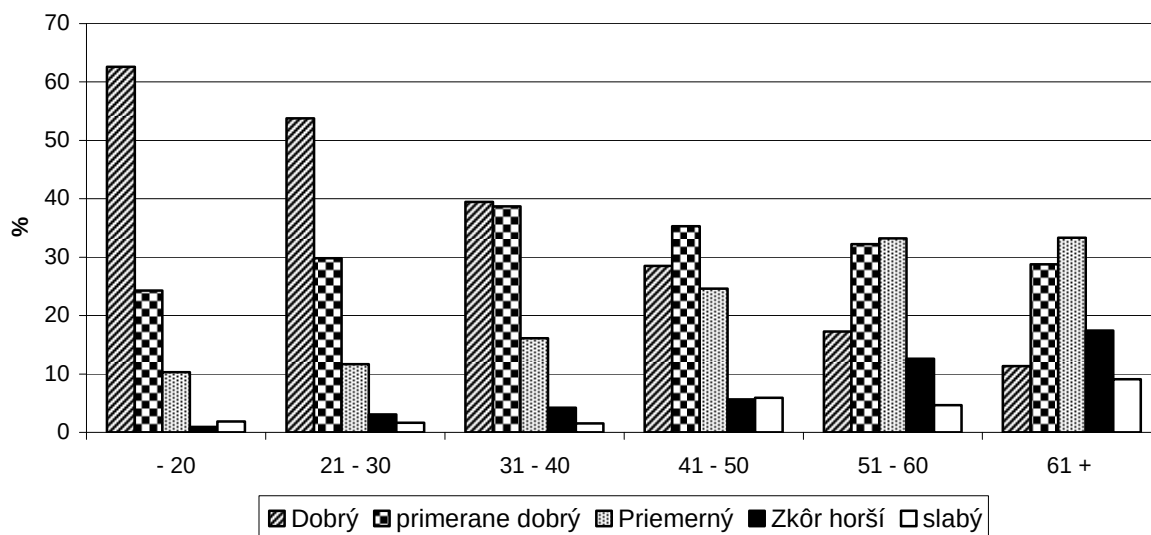


Zdroj: CINDY.

G r a f 5 Hodnotenie zdravotného stavu v populácii Slovenska

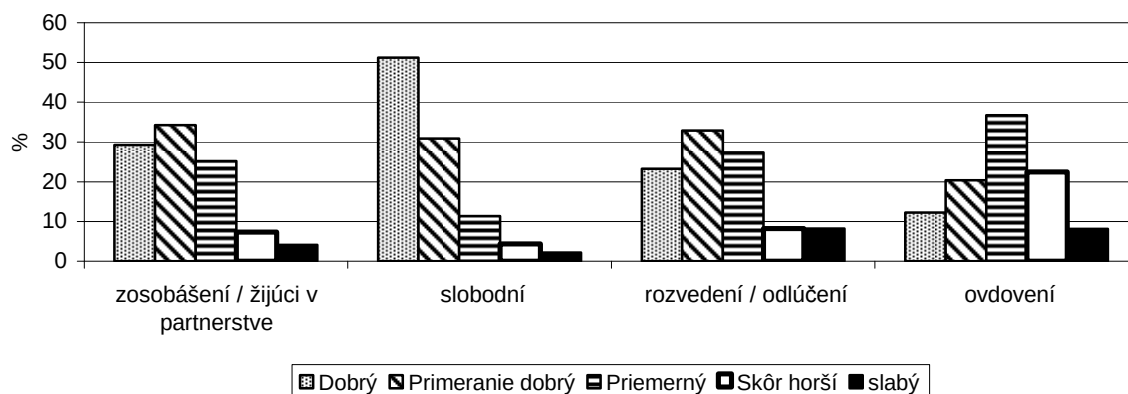
Hodnotenie zdravotného stavu podľa vekových skupín potvrdilo naše očakávania. Pozitívne hodnotenie klesá s rastom veku (Graf 6). Viac ako 60 % populácie mladšej ako 20 rokov hodnotí svoje zdravie ako dobré. Vo vekovej skupine 61 a viac ročných hodnotí svoj zdravotný stav ako dobrý mierne nad 10 %.

Nepriamym dôkazom závislosti samohodnotenia od veku je rozdelenie podľa rodinného stavu respondentov (Graf 7). Najvyšší podiel obyvateľov pozitívne hodnotiacich svoj zdravotný stav je v skupine slobodných, naopak v skupine ovdovených prevládajú negatívnejšie hodnotenia. Je vysoko pravdepodobné, že v skupine slobodných prevládajú mladí ľudia a v skupine ovdovených sa nachádzajú prevažne staršie ročníky. Zreteľný je rozdiel medzi skupinou zosobášených alebo žijúcich v partnerstve oproti rozvedeným alebo odlúčeným. Hlavným faktorom je rozdielna veková štruktúra týchto dvoch skupín (Graf 8).

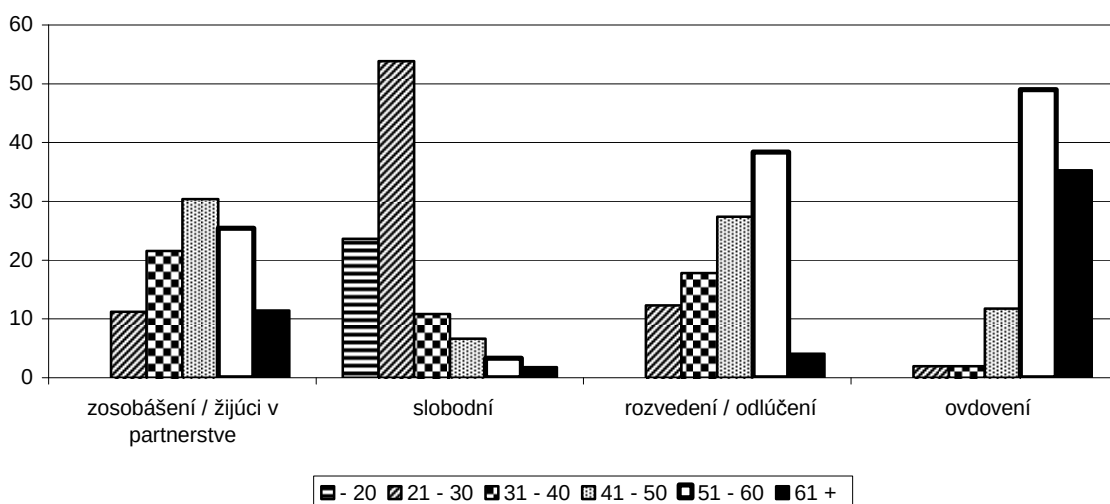


Zdroj: CINDY.

G r a f 6 Hodnotenie zdravotného stavu podľa veku



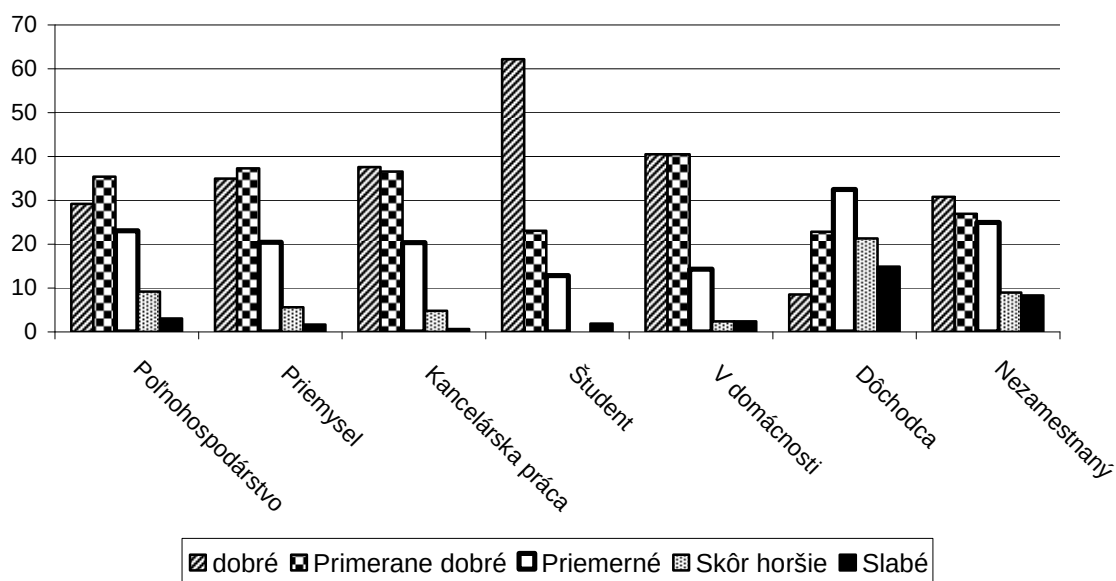
Zdroj: CINDY.

G r a f 7 Hodnotenie zdravotného stavu podľa rodinného stavu

Zdroj: CINDY.

G r a f 8 Rodinný stav vo vekových skupinách

Vzhľadom na druh ekonomickej aktivity majú najlepšie hodnotený stav študenti. Naopak najhoršie hodnotenia sú v skupine dôchodcov a nezamestnaných. Príčiny sú zrejmé. Študenti reprezentujú najmladšiu vekovú skupinu a dôchodcovia najstaršiu. Medzi nezamestnanými môže zohrávať úlohu životná úroveň a sklon k negatívnejšiemu hodnoteniu svojho stavu. Spomedzi skupín zamestnaných majú vyššiu frekvenciu hlásenia dobrého zdravotného stavu kancelársky pracovníci, ako pracovníci v priemysle či poľnohospodárstve. Hlavným dôvodom môže byť vplyv zdravotného stavu na ich pracovné výkony. Kancelárskym pracovníkom častokrát dokážu „pretrpieť“ menšie zdravotné problémy, ktoré ich výraznejšie neobmedzujú v pracovných výkonoch.



Zdroj: CINDY.

G r a f 9 Hodnotenie zdravotného stavu podľa ekonomickej aktivity

4 Vplyv pohlavia a ďalších faktorov na zdravotný stav a využitie zdravotných služieb (logit analýza)

Pre exaktnejšie hodnotenie vplyvu jednotlivých faktorov na zdravotný stav a využitie zdravotných služieb bol skonštruovaný model typu logit. Boli vypočítané dva základné modely. Prvý model popisuje samohodnotenie zdravotného stavu a druhý popisuje využitie zdravotných služieb. Vysvetľujúce umelé premenné popisuje tabuľka 1.

T a b u ľ k a 1 Kódovanie umelých nezávislých premenných

Premenná	Kategória	Kódovanie					
		1	2	3	4	5	6
POVOLANIE	POĽNOHOSPODÁRSTVO	1	0	0	0	0	0
	PRIEMYSEL	0	1	0	0	0	0
	KANCELÁRSKA PRÁCA	0	0	1	0	0	0
	ŠTUDENT	0	0	0	1	0	0
	V DOMÁCNOSTI	0	0	0	0	1	0
	DÔCHODCA	0	0	0	0	0	1
	NEZAMESTNANÝ	0	0	0	0	0	0
RODINNÝ STAV	SLOBODNÝ	0	0	0			
	SOBÁŠ ALEBO PARTNER	1	0	0			
	ROZVEDENÝ	0	1	0			
	OVDOVENÝ	0	0	1			
VZDELANIE	ZÁKLADNÉ	0	0	0			
	NIŽŠIE STREDNÉ	1	0	0			
	STREDNÉ	0	1	0			
	UNIVERZITA	0	0	1			
ROD	ŽENA	1					
	MUŽ	0					
VEKOVÁ SKUPINA	MLADÝ	0					
	STARÝ	1					

5 Faktory ovplyvňujúce dobré zdravie

V analýze bol vyhodnotený vplyv dobrého samohodnoteného zdravotného stavu. Ako závislá premenná boli vytvorené dve umelé premenné 0 – zle hodnotený zdravotný stav (kategórie skôr horšie, slabé) a 1 – dobre hodnotený stav (kategórie veľmi dobré, dobré, priemerné). Ako vysvetľujúce premenné boli použité premenné z tabuľky 1. Skrátené výsledky sú prezentované v tabuľke 2.

Modelom bolo dokázané, že neexistuje štatistická závislosť medzi hodnotením zdravotného stavu a pohlavím. Preukazným sa ukázal vplyv veku. Premenná STARÝ prezentuje vekovú skupinu 41 a viac. Vzhľadom na ekonomickú aktivitu študenti hodnotia svoj zdravotný stav najlepšie a dôchodcovia najhoršie. Zamestnaní v kancelárii hodnotia zdravotný stav lepšie ako pracovníci v priemysle a v poľnohospodárstve. Príčiny môžu byť v rozdielnych typoch práce. Pracovníci v kancelárii pracovne nezaťažujú svoj organizmus tak pracovníci v priemysle a poľnohospodárstve. Vyššie vzdelanie zvykne byť spojené s lepšími zdravotnými návykmi.

T a b u ľ k a 2. Výsledky logit regresie zdravotného stavu

Premenná	Koef.	Pravdep.
C	2,040	0,000
Žena	0,046	0,812
Starý	-0,600	0,057
Poľnohosp.	0,429	0,340
Priemysel	0,873	0,007
Kancelária	0,902	0,004
Študent	2,214	0,001
V domácnosti	1,320	0,086
Dôchodca	-0,924	0,001
Nižšie stredné	-0,744	0,004
Stredné	0,334	0,148
Univerzita	0,545	0,094
Zosobášený	0,107	0,702
Rozídený	-0,236	0,580
Ovdovený	-0,138	0,759

Pri porovnaní úrovne vzdelania, s vyšším vzdelaním rastie frekvencia hodnotenia zdravotného stavu ako lepšieho. Čím vyššie vzdelanie, tým lepšie je hodnotený zdravotný stav. Príčina je pochopiteľná. Vzdelaní ľudia majú odlišný životný štýl v porovnaní s nízko vzdelanými. Lepší postoj k svojmu zdraviu je jedným z podstatných rozdielov. Môže to byť spojené samozrejme s rozšírenejším príjmom zdravšieho jedla a väčšími možnosťami uspokojenia osobných a rodinných potrieb prosievajúcich zdraviu, napr. voľnočasových aktivít. Treba vidieť takisto vzťah s lepšími zdrojmi informácií a celkovým lepším zdravotným stavom. Rodinný stav nemá žiaden vplyv na hodnotenie zdravotného stavu.

6 Faktory ovplyvňujúce využitie zdravotných služieb

Logit model využitia zdravotnej starostlivosti bol založený na rovnakých nezávislých premenných ako predošlá analýza hodnotenia zdravotného stavu. Do modelu bude zaradená ešte jedna premenná DOBRÝ, ktorá popisuje respondentov, ktorí hodnotili svoj zdravotný stav ako veľmi dobrý, dobrý alebo priemerný. To by nám malo pomôcť identifikovať, či občania s dobrým zdravotným stavom majú rovnaký dopyt po zdravotných službách ako tí, ktorí označujú svoje zdravie ako zlé. Závislou premennou bola frekvencia návštev u lekára.

Tabuľka 3 Výsledky regresie využitia zdravotnej starostlivosti

Premenná	Koef.	Pravdep.
C	-0,601	0,250
Žena	0,291	0,078
Starý	-0,046	0,843
Poľnohosp,	0,709	0,069
Priemysel	0,887	0,001
Kancelária	0,611	0,023
Študent	0,713	0,042
V domácnosti	0,758	0,108
Dôchodca	-0,090	0,817
Nižšie stredné	0,416	0,147
Stredné	0,290	0,131
Univerzita	0,252	0,303
Zosobášeny	-0,126	0,564
Rozídený	0,089	0,817
Ovdovený	-0,109	0,840
Dobrý	0,273	0,567

Pri pohľade na výsledky (Tabuľka 3) je možné vidieť, že sú u analyzovaných skupín niektoré povšimnutia hodné rozdiely.

7 Záver

Z uvedenej analýzy determinantov využívania zdravotných služieb vyplýva, že ženy využívajú zdravotnú starostlivosť viac ako muži. Tento záver je možné urobiť iba v prípade posunutia hladiny významnosti na 0,1.² To znamená, že závery by týmto smerom mali byť brané s istou mierou opatrnosti. To korešponduje s výsledkami priemerného využitia zdravotných služieb uvedených vyššie. Model nepreukázal vplyv veku na využitie zdravotných služieb. Koeficient je jednak nevýznamný a jednak má opačné znamienko ako sme očakávali v predpokladoch. Zaujímavé sú výsledky využitia zdravotných služieb na základe ekonomickej aktivity. Bola zistená pozitívna závislosť v spotrebe medzi pracovníkmi v priemysle, kancelárii a študentmi. Naproti tomu vyššie využitie medzi dôchodcami nie je preukázateľné. Takisto nie je preukázaný vplyv vzdelania a využitia zdravotných služieb. Rovnako ani hodnotenie zdravotného stavu neprinieslo očakávané výsledky.

² Pre väčšinu analýz bola použitá hladina významnosti 0,05.

8 Literatúra

- [1] AVDIČOVÁ, M., F. HRUBÁ, O. HERDOVÁ, J. NÁMEŠOVÁ AND K. FRANCISCOVÁ (2003), Health Habits and the Quality of the Adult Population Health in Slovakia in the Year 2002, State Health Institute, Banská Bystrica.
- [2] HAULÍKOVÁ, L. (2000), "Financing of social security", in A. Marcinčin and M. Beblaný (eds), Economic policy in Slovakia 1990–1999, INEKO, Bratislava.
- [3] Institute of Public Health (1992–2001), Health Awareness and Behaviour of the Population in the Slovak Republic, Bratislava.
- [4] Ministry of Finance (2002), Analysis of financial resources for healthcare and their proportion of GDP (sources from the Ministry of Economy of SR and sources from health insurance companies), Bratislava.
- [5] PÁLENÍK, V., K. KRIVANSKÁ, V. KVETAN, A. PAULOVIČ, G.G. ROBERTS AND J. ŠIMO (2002), Mid-term projections of trends in the macroeconomic development of the Slovak health sector, EMPO, Bratislava.
- [6] PAŽITNÝ, P. AND R. ZAJAC (2001), "Healthcare reform – Reform for citizens", *MESA* 10, Bratislava, February.
- [7] UZIŠ [Institute of Health Information and Statistics] (1996–2002), Health Statistical Yearbook 1996–2002, Bratislava.
- [8] World Health Organisation (WHO) (2000), CINDI Health Monitor Questionnaire, WHO Regional Office for Europe, Copenhagen.
- [9] VLADIMÍR KVETAN, VILIAM PÁLENÍK, MARTIN MLÝNEK, MAREK RADVANSKÝ: Starnutie, zdravotný stav a determinanty výdavkov na zdravie v podmienkach Slovenska. Ekonomický ústav SAV VS 1, Bratislava 2007, ISBN 978-80-7144-160-1
- [10] AHEAD Studies Published in the ENEPRI Research Reports Series (available for free downloading from www.enepri.org or <http://shop.ceps.be>)
- [11] No. 30 Health and Morbidity in the Accession Countries: Country Report – Slovakia, Vladimír Kvetan, Viliam Páleník, December 2006
- [12] No. 31 Health Status and Healthcare Systems in Central and East European Countries: Bulgaria, Estonia, Poland, Slovakia and Hungary, December 2006

Adresa autorov:

Vladimír Kvetan, Ing.
Ekonomický ústav SAV
Šancová 56
811 01 Bratislava
vladimir.kvetan@gmail.com

Viliam Páleník, h.doc. RNDr., PhD.
Ekonomický ústav SAV
Šancová 56
811 01 Bratislava
viliam.palenik@savba.sk

Dlhodobá starostlivosť na Slovensku

Long-term care in Slovakia

Marek Radvanský

Abstract: This paper is covering and describing basic principles of Long Term Care (LTC) in Slovakia, analysing future needs of LTC and presenting actual problems in Slovakia. It brings brief information about available LTC data, statistics and current situation in European Union.

Key words: Long-term care, ageing, funding scheme

Kľúčové slová: dlhodobá starostlivosť, starnutie, schéma financovania

Úvod

Dlhodobá starostlivosť (LTC – Long Term Care) je jednou z oblastí, ktorej je na Slovensku stále venovaná malá pozornosť. Starnutie populácie je fenoménom, ktorý sa v nasledujúcich rokoch silno prejaví takmer vo všetkých krajinách Európskej únie. Zo starnutím obyvateľstva narastá pravdepodobnosť a počet obyvateľov odkázaných na určitú formu starostlivosti. Slovenská republika zareagovala na problém starnutia aj v Programovom vyhlásení vlády prijatím Národného programu starostlivosti o seniorov. Napriek tomu, táto problematika je i naďalej zanedbávaná.

Systém dlhodobej starostlivosti je systém poskytovania komplexných zdravotných, ošetrovateľských a opatrovateľských služieb v dlhšej časovej perióde (zvyčajne sa chápe minimálna hranica očakávanej starostlivosti v rozsahu 6-12 mesiacov) a trvá v niektorých prípadoch doživotne. Širšia definícia systému LTC zahŕňa zdravotné a sociálne poradenstvo, služieb podpory integrácie, kompenzačných pomôcok a peňažných príspevkov v takom rozsahu, aby bola zmiernená, alebo nahradená strata schopnosti vykonávať základné každodenné sociálne a pracovné funkcie.

Tento príspevok vznikol na Ekonomickom Ústave SAV, ako súčasť riešenia výskumného projektu ANCIEN (Assessment of Needs of Care in European Nations) v rámci 7. rámcového programu únie. Tento projekt sa venuje odhadu potrieb dlhodobej starostlivosti v členských štátoch EÚ. Zameriava sa na odhad budúcich potrieb jednak z dopytovej stránky, teda odhadu počtu ľudí vyžadujúcich určitú formu LTC a zároveň odhady budúcich možností poskytovania týchto služieb (vrátane domácej ošetrovateľskej a opatrovateľskej starostlivosti). Zameranie zahŕňa jednak poskytovanie inštitucionálnej dlhodobej starostlivosti, tak aj domácej či ambulantnej formálnej a neformálnej starostlivosti. Tento príspevok sa venuje popisu základných determinantov dlhodobej starostlivosti na Slovensku v porovnaní s krajinami EÚ.

Budúci vývoj systému dlhodobej starostlivosti na Slovensku bude podmienený celkovým vývojom obyvateľstva a vývojom zdravotného stavu populácie. Napriek tomu tento systém nie je dostatočne podchytený štatistickými dátami, pretože systém LTC nie je integrovaný a prelína sa v rámci systémov zdravotnej a sociálnej starostlivosti, ktoré si vypracovávajú a poskytujú samostatné štatistiky. Čiastočnou náhradou môžu byť úzko

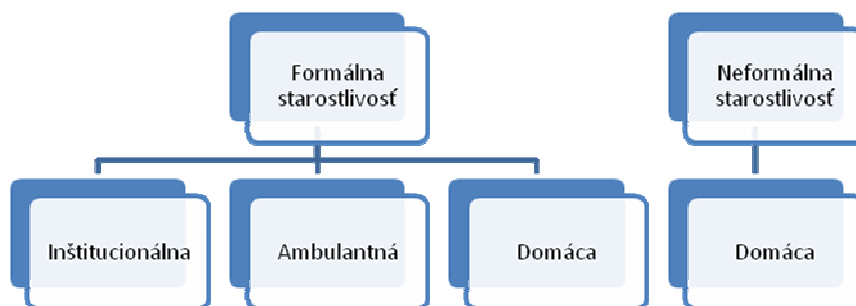
zamerané štatistické zisťovania. Významným zdrojom dát je európske štatistické zisťovanie EUROBAROMETER 67.3, ktoré čiastočne zahŕňa aj otázky týkajúce sa dlhodobej starostlivosti.

Systém LTC na Slovensku

Integrácia sociálnej a zdravotnej starostlivosti nie je systematicky regulovaná národnou legislatívou. Napriek tomu, prístup k systému dlhodobej starostlivosti na Slovensku za posledných 15 rokov prešiel niekoľkými výraznými zmenami. V roku 2005 bol prezentovaný návrh zákona o *dlhodobej starostlivosti a dlhodobej podpore integrácie osôb s funkčným obmedzením* navrhnutý ministrom zdravotníctva Rudolfom Zajacom. Tento materiál komplexne riešil a navrhoval integrovaný systém dlhodobej starostlivosti, legislatívny proces však bol zastavený.

V súčasnosti sú jednotlivé časti systému dlhodobej starostlivosti regulované viacerými zákonmi. Dlhodobá starostlivosť je rozčlenená na dve oblasti podľa typu potrieb – na sociálnu a zdravotnú. Táto oblasť zahŕňa aj dlhodobú starostlivosť postihnuté osoby a problém ich potrieb a sociálnej inklúzie. Ďalšia časť je venovaná geriatrickej starostlivosti a osobám z chronickými chorobami. Hlavné princípy sú definované v niekoľkých národných sociálnych, zdravotných stratégiách a zákonov pokrývajúcich túto problematiku (hlavne zákony 576/2004, 578/2004, 447/2008 a 448/2008)

Graf 1 – Prehľad systému LTC na Slovensku



V roku 2006 bol podiel osôb starších ako 65 na populácii 11,8 %, čo predstavovalo približne 640 tisíc osôb. Podiel ľudí starších ako 60 rokov bol viac ako 814 tisíc, čo reprezentovalo 15 % obyvateľov. Podľa Ministerstva zdravotníctva 38 % populácie dôchodcov je relatívne zdravá, 32 % má nejaký druh chronického ochorenia a je im poskytovaná zdravotná starostlivosť. Zostávajúcich 21 % je ohrozených orgánovou dekompenzáciou a 9 % má závažnú chorobu alebo je závislých na dlhodobej starostlivosti poskytovanej treťou osobou.

Potreba dlhodobej starostlivosti (vrátane demografických charakteristík)

Slovenská legislatíva zaručuje univerzálny prístup k službám dlhodobej starostlivosti pre každú osobu, ktorá ju potrebuje. Slovensko má podobne ako väčšina krajín problémy s narastajúcim podielom starších osôb v populácii – so starnutím. Očakávaný demografický vývoj hlavných vekových skupín je zobrazený v tabuľke 1.

Výrazný pokles obyvateľov v predproduktívnom veku sa očakáva hlavne do roku 2015 a dosiahne minimum približne okolo 15 % celkovej populácie Slovenska. V tomto období zároveň mierne vzrastie skupina obyvateľov v produktívnom veku z 56% v roku 2015 na takmer 62 % v roku 2025. Takmer kritická situácia nastane vo vekovej skupine 65+, keď podiel tejto skupiny vzrastie z 50% v 2005 na 62% počas ďalších 20 rokov. Podiel veľmi starých ľudí sa v tomto období zdvojnásobí.

Podľa odhadu slovenského Ministerstva práce, sociálnych vecí a rodiny potrebuje dlhodobú starostlivosť približne 183 tisíc ľudí a takmer viac ako polovica z nich má viac ako 65 rokov. Viac ako 38 tisíc prijíma formálnu ústavnú starostlivosť, viac ako 70 tisíc poberá nejaký typ formálnej domácej starostlivosti a viac ako 60 tisíc je prijímateľom určitého typu neformálnej starostlivosti. Hrubý odhad hovorí, že približne 20 % osôb, ktoré starostlivosť potrebujú, nedostávajú žiadny typ dlhodobej starostlivosti.

Na základe Správy EK o starnutí z roku 2009 je odhad počtu osôb, ktoré potrebujú dlhodobú starostlivosť ešte vyšší (približne 230 tisíc osôb, z ktorých viac ako 31 tisíc prijíma formálnu ústavnú starostlivosť). Tento počet sa pravdepodobne v najbližších 25 rokoch viac ako zdvojnásobí.

Tabuľka 1 – Demografická prognóza vybraných vekových skupín

	2000	2005	2010	2015	2020	2025
Populácia	5 400 679	5 387 285	5 423 703	5 471 653	5 510 225	5 521 745
Predproduktívny vek (0-14)	1 210 798	1 053 386	906 823	820 187	827 994	853 365
Produktívny vek (15-64)	2 580 223	2 684 915	2 864 970	3 077 306	3 265 048	3 408 341
Poproduktívny vek (65+)	617 516	630 927	675 883	775 472	921 798	1 047 470
Veľmi staré osoby(80+)	102 737	130 609	153 337	167 911	179 048	202 109

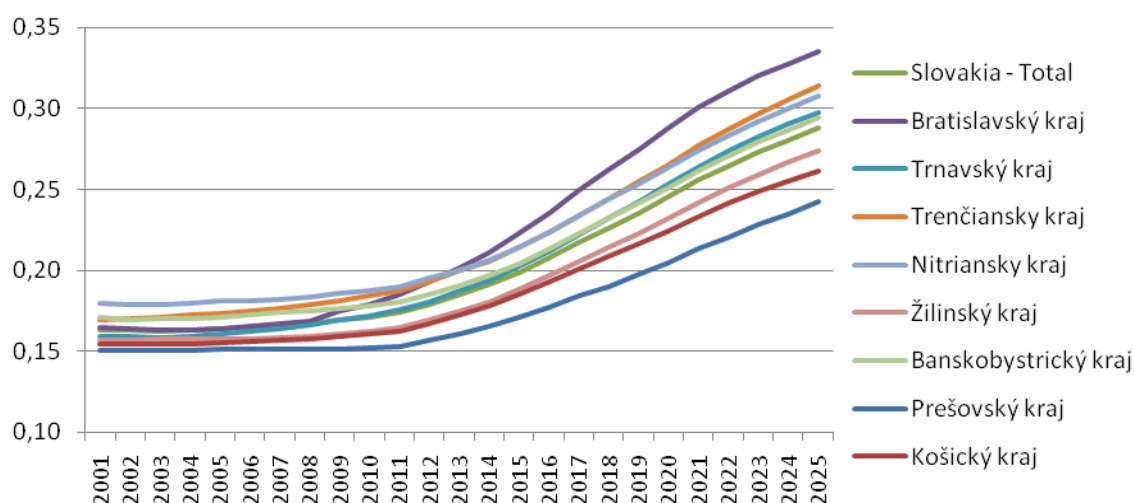
Zdroj: Výskumné demografické centrum, 2008

Tabuľka 2 – Demografická prognóza vybraných vekových skupín, podiel na celkovej populácii

	2000	2005	2010	2015	2020	2025
Predproduktívny vek (0-14)	22,4 %	19,6 %	16,7 %	15,0 %	15,0 %	15,5 %
Produktívny vek (15-64)	47,8 %	49,8 %	52,8 %	56,2 %	59,3 %	61,7 %
Poproduktívny vek (65+)	11,4 %	11,7 %	12,5 %	14,2 %	16,7 %	19,0 %
Veľmi staré osoby (80+)	1,9 %	2,4 %	2,8 %	3,1 %	3,2 %	3,7 %

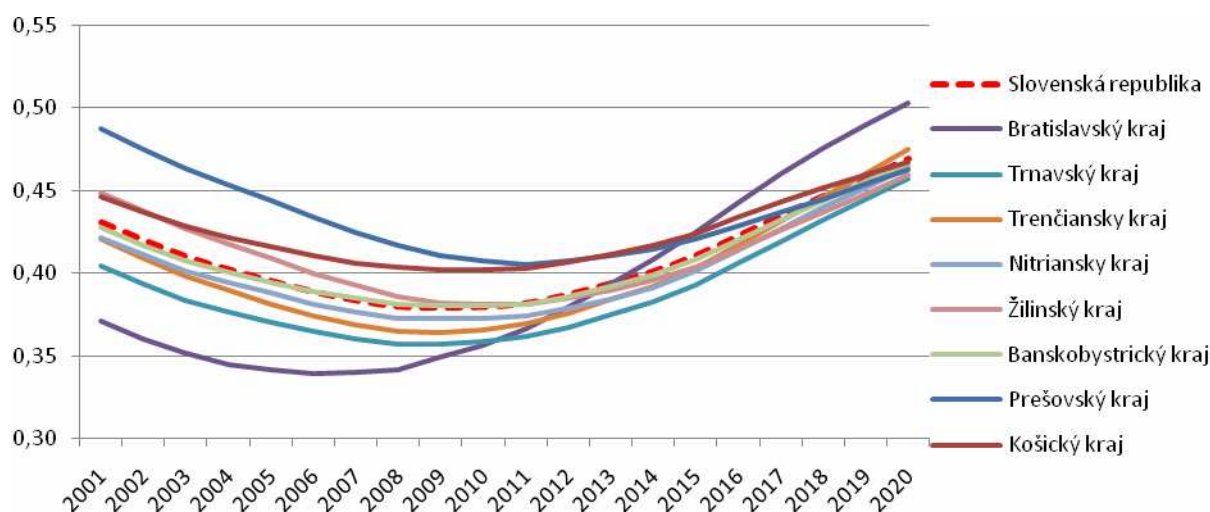
Zdroj: Výskumné demografické centrum, 2008

Index ekonomickej závislosti starých ľudí charakterizuje vývoj podielu populácie staršej ako 64 na populácii v produktívnom veku (15-64). Do roku 2008 rástol tento ukazovateľ veľmi pomaly. Od roku 2008 staršie osoby narodené počas povojnového baby boom-u, predstavujúce veľmi silné kohorty, odídu do dôchodku. Závislosť starých ľudí bude v blízkej budúcnosti rásť extrémne rýchlo a za 15 rokov sa zdvojnásobí. Najvyšší bude index závislosti v ekonomicky najsilnejšej časti západnej časti Slovenska (Graf 2).

Graf 2 – Index ekonomickej závislosti starých ľudí, slovenské regióny NUTS 3

Zdroj: Výskumné demografické centrum a vlastné prepočty

Ešte výraznejšie problém starnutia ilustruje vývoj indexu ekonomického zaťaženia. Index ekonomického zaťaženia vyjadruje, aký je počet ľudí v predproduktívnom a poproduktívnom veku na jedného človeka v produktívnom veku. Do roku 2009 hodnota tohto indikátora klesala (Graf 3). Od roku 2010 sa očakáva že ekonomické zaťaženie bude rásť v priemere medziročne za všetky kraje 2,1 %. Najvyšší nárast bude vykazovať Bratislavský kraj na úrovni 3,4 % medziročne. Kým v minulosti bol najväčší počet ekonomicky závislých ľudí v predproduktívnom veku (deti), v súčasnosti sa toto zaťaženie presúva vplyvom rastúceho počtu starších ľudí. Momentálne index ekonomického zaťaženia dosahuje minimálne hodnoty na priemernej úrovni pod 0,38.

Graf 3 - Index ekonomického zaťaženia v rokoch 2001-2020

Zdroj: ŠÚSR, VDC a prognózy EÚ SAV

Dopyt a ponuka neformálnej starostlivosti

Neformálna starostlivosť je najčastejšie využívaná pomoc v systéme dlhodobej starostlivosti. Takmer každá osoba odkázaná na pomoc prijíma nejakú formu neformálnej starostlivosti. Rôzne zdroje uvádzajú, že na Slovensku je približne 200 tisíc osôb, ktoré potrebujú dlhodobú starostlivosť, čo predstavuje približne 3,6 % celkovej populácie. Intenzívna neformálna starostlivosť je poskytovaná približne 57 tisícom ľudí.

Dopyt a ponuka formálnej starostlivosti

Exaktné dáta o počte osôb dopytujúcich dlhodobú starostlivosť neexistujú. Odhad založený na počte ľudí žiadajúcich akýkoľvek druh starostlivosti je približne 183 tisíc ľudí. V rámci sústavného starnutia populácia očakávame signifikantný nárast v dopyte po dlhodobej starostlivosti. V nasledujúcich 10 rokoch by nárast v dopyte po dlhodobej starostlivosti spôsobený starnutím populácie mohol predstavovať 50 %.

Odhadovaný počet zamestnancov v systéme slovenskej dlhodobej starostlivosti je približne 32 tisíc. Približne 22 tisíc pracuje v systéme ústavnej starostlivosti a viac ako 10 tisíc predstavujú ošetrovatelia alebo zdravotné sestry v domácej starostlivosti.

Tabuľka 3 – Počty a miesta vo vybraných sociálnych zariadeniach

	2001	2002	2003	2004	2005	2006	2007	2008
Celkový počet zariadení	670	590	680	730	775	797	824	873
Počet miest v domovoch dôchodcov	13374	12922	12666	13214	13277	13258	13758	13922
Počet miest v domovoch sociálnych služieb pre hendikepovaných dospelých	6569	8330	9627	9902	10839	12444	12833	13249
Počet miest v domovoch sociálnych služieb pre hendikepované deti	3668	3749	3636	3659	2961	2345	2088	2371

Zdroj: Štatistický úrad SR

Ústavná starostlivosť

Ústavná starostlivosť na Slovensku tvorí viacero typov zariadení. V zariadeniach celého systému dlhodobej starostlivosti na Slovensku je viac ako 38 tisíc postelí. Tento počet nie je dostatočný najmä v oblasti dlhodobej starostlivosti o starších ľudí. V niektorých zariadeniach sú čakacie lehoty aj niekoľko rokov. Viac ako 2/3 všetkých postelí patrí do štátnych zariadení, zvyšné sú súkromné. Väčšina súkromných zariadení bola založená na poskytovaní rôznych druhov sociálnej starostlivosti.

Domáca starostlivosť

Domáca starostlivosť je prioritou pre ľudí zo zaisteným sociálnym pozadím a je najmenej stresujúca. Najdôležitejšia časť domácej starostlivosti je neformálna starostlivosť poskytovaná rodinou, ktorá je poskytovaná všetkým osobám vyžadujúcim nejaký druh dlhodobej starostlivosti. Neformálna starostlivosť je poskytovaná rodinou alebo blízkymi osobami. V niektorých prípadoch môže byť takáto starostlivosť financovaná štátnymi príspevkami v zmysle náhrady príjmu. Druhým typom je formálna domáca starostlivosť a formálne domáce ošetrovanie. V roku 2007 bolo približne 70 tisíc prijímateľov jedného

z týchto druhov formálnej domácej starostlivosti, pričom z tohto počtu bolo viac ako 40 tisíc osôb starších ako 65 rokov.

Semi-ústavná starostlivosť

Menšiu časť systému dlhodobej starostlivosti predstavujú inštitúcie poskytujúce najmä dočasnú starostlivosť založenú na dennej báze, ako denné stacionár, mobilný hospic a ambulantná starostlivosť.

Financovanie systému LTC

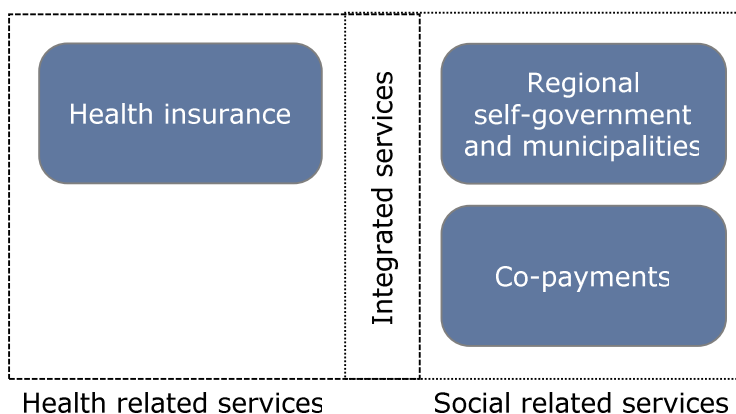
Systém financovania dlhodobej starostlivosti na Slovensku je zmiešaný. Výdavky na dlhodobú starostlivosť sú financované z dvoch zdrojov, ktoré sa odlišujú podľa druhu poskytovanie dlhodobej starostlivosti. Služby spojené so starostlivosťou o zdravie sú plne hrazené zo zdravotného poistenia, formálne neexistuje žiadna spoluúčasť zo strany prijímateľa. Sociálna starostlivosť je financovaná verejnou a regionálnou správou prostredníctvom daní. Toto pokrýva približne 2/3 výdavkov na sociálnu starostlivosť. Zvyšná 1/3 zostáva vo forme spoluúčasti pacientov na starostlivosť. Spadá sem ústavná aj domáca starostlivosť, pričom ošetrovanie zdravotnou sestrou v domácom prostredí tiež nevyžaduje platby zo strany prijímateľa.

Zdravotné a sociálne poistenie sú povinné. Odvodové platby sú rozdelené medzi zamestnanca a zamestnávateľa. Štát platí tieto príspevky za deti do ukončenia školy, rodiča starajúceho sa o dieťa do veku 6 rokov, starobných a invalidných dôchodcov a osoby poskytujúce domácu starostlivosť pacientom, ktorí sú na ňu odkázaní.

Inštitúcia poskytujúca sociálnu starostlivosť (súkromnú alebo verejnú) dostáva príspevok za každého pacienta na základe jeho stupňa odkázanosti na takúto pomoc. Súkromný poskytovateľ je spôsobilý na takéto príspevky len pokiaľ má zmluvu s lokálnou a/alebo regionálnou správou. Každý pacient platí za služby, pričom jeho podiel na nákladoch závisí od jeho príjmu. Výška platieb mu musí dovoliť ponechať si minimálne 1,3-násobok nákladov na živobytie (životné minimum).

Celkové náklady na dlhodobú starostlivosť na Slovensku sú približne 130 mil. Eur, čo predstavuje približne 0,2 % HDP. Detailnejšie rozpracované finančné štatistiky týkajúce sa systému dlhodobej starostlivosti na Slovensku chýbajú.

Graf 4 – Schéma financovania Slovenského systému dlhodobej starostlivosti

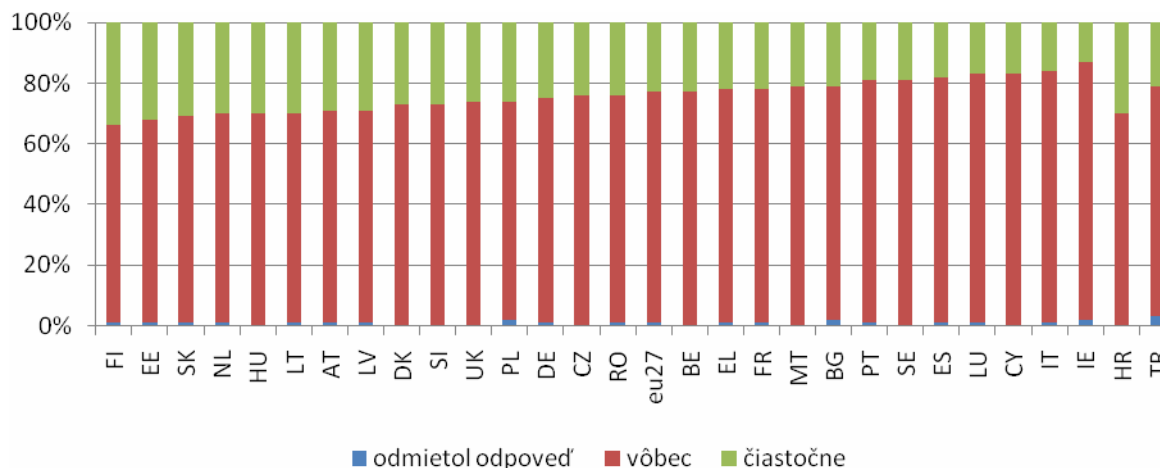


Eurobarometer 67.3

Čiastkové odpovede o zdravotnom stave obyvateľstva a stave dlhodobej starostlivosti z pohľadu obyvateľstva sa môžeme dozvedieť z výberových zisťovaní. Niektoré z nich, ako napr. Eurobarometer sú robené na Európskej úrovni a dotazník je rovnaký pre všetky krajiny, čo je kľúčové pre porovnateľnosť dát. Aj napriek tomu je nutné uvedomiť si niektoré limitácie takéhoto prieskumu, napr. rôzne pochopenie otázky v rôznych krajinách, ak je systém poskytovania starostlivosti mierne odlišný, či celkovú mentalitu a spokojnosť populácie.

V tejto časti budú uvedené príklady otázok a ich vyhodnotenia v rámci krajín EÚ.

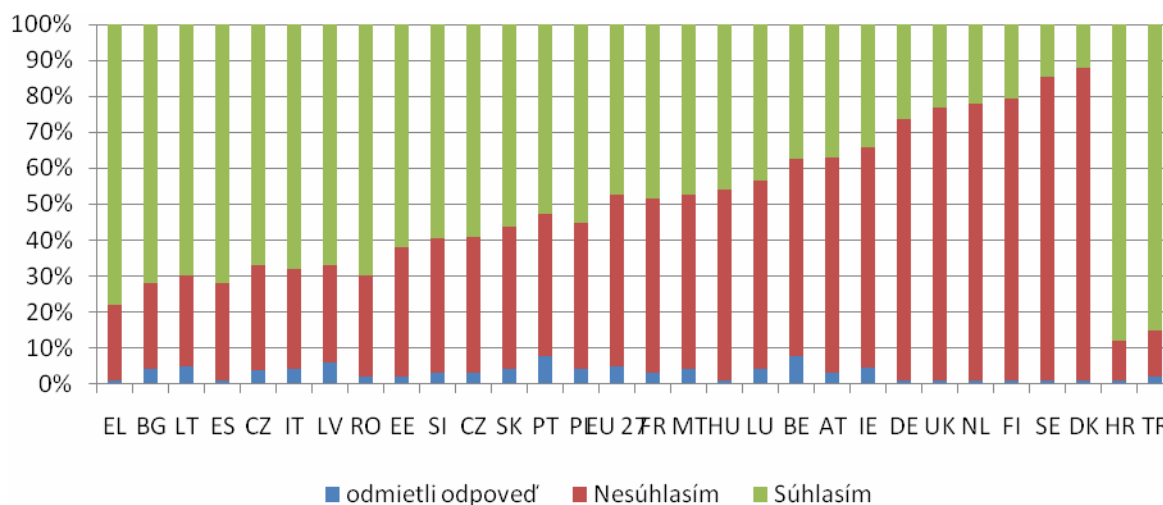
Graf 5 – Podiel osôb nejakým spôsobom limitovaných svojim zdravotným/psychickým stavom za posledných 6 mesiacov



Zdroj: Eurobarometer

Prieskumy umožňujú odhaliť zmýšľanie obyvateľstva a špecifiká niektorých krajín. Dobré to ilustruje odpoveď na otázku o platení detí za starostlivosť o svojich rodičov. V krajinách z veľmi silným sociálnym systémom (hlavne severské krajiny) je pre každého obyvateľa jasné, že ich sociálny systém s vysokým prerozdeľovaním sa o ich rodičov bez problémov postará. Opačná situácia je napríklad v Turecku. Pri krajinách zo spoločenstva pacienta obyvateľia predpokladajú čiastočnú pomoc od ich detí.

Graf 6 – Vyhodnotenie odpovedi na otázku: Majú deti platiť za ošetrovanie ich rodičov, pokiaľ to ich príjem nedovoľuje?



Zdroj: Eurobarometer

Kritické posúdenie systému dlhodobej starostlivosti

Súčasná situácia v systéme dlhodobej starostlivosti na Slovensku nie je veľmi transparentná. Je to tiež jeden z hlavných problémov z legislatívneho hľadiska. Jednotlivé súčasti súvisiace so zdravotnou a sociálnou starostlivosťou sú ošetrované rôznymi zákonmi a reguláciami, ktoré nie sú vždy prepojené (integrované) a niekedy nepokrývajú celú súvisiacu oblasť. V minulosti bolo predstavených niekoľko koncepcií a návrhov. V rámci reformy zdravotníctva bol pripravený aj návrh zákona o dlhodobej starostlivosti, ale stále nebol schválený z politických dôvodov.

Problémy starnutia obyvateľstva a zvyšujúci sa dopyt po službách dlhodobej starostlivosti v poslednej dobe pútajú pozornosť a stávajú sa kritickým problémom spoločnosti. Budúca konečná podoba systému dlhodobej starostlivosti je stále sporná. Dnes môžeme povedať, že jednotlivé časti systému dlhodobej starostlivosti na Slovensku, ktoré sú chápané ako rámec národných stratégií, majú niekoľko definícií a plánov, ktoré sú podobné návrhom systému dlhodobej starostlivosti svetovej zdravotníckej organizácie - WHO. Z nášho uhla pohľadu je najväčším súčasným problémom riešenie neformálnych služieb dlhodobej starostlivosti, ktoré nie sú legislatívou komplexne riešené vôbec.

Zákon č. 448/2008 o sociálnych službách, platný od januára 2009, dovoľuje spájanie sociálnej a medicínskej starostlivosti rovnakým spôsobom, aký bol v platnosti pred 15 rokmi. Najdôležitejším faktom je, že sestry zamestnané v sociálnych zariadeniach môžu poskytovať medicínsku starostlivosť. Do decembra 2008 bolo nutné, aby klientom sociálnych služieb poskytovala medicínsku starostlivosť zdravotná sestra privolaná z medicínskeho zariadenia.

V Slovenskom systéme dlhodobej starostlivosti sa vyskytlo niekoľko bariér. Sú to najmä finančné bariéry, geografické bariéry (nevybilancovaný rozdelenie poskytovateľov) a organizačné bariéry (niekedy sú čakacie lehoty na sociálnu starostlivosť pre starších ľudí aj niekoľko rokov). Systém rôznych druhov dlhodobej starostlivosti je taktiež dosť široký, je potrebné ho zjednodušiť.

Služby dlhodobej starostlivosti sa nachádzajú v tranzitívnej fáze. Prebieha definícia budúcich potrieb, bude dôležité špecifikovať úlohy a priority v poskytovaní neformálnej starostlivosti. Z hľadiska ohodnotenia existujú rozličné názory na budúce potreby a na

optimálnu skladbu dlhodobej starostlivosti zloženú z poskytovateľov na národnej, regionálnej a lokálnej úrovni.

Záver

V článku sme poukázali na vybrané problémy týkajúce sa systému dlhodobej starostlivosti na Slovensku. Predložené časti iba veľmi stručne ilustrujú fungovanie systému dlhodobej starostlivosti na Slovensku. Problematika starnutia a s ňou spojený zvyšujúci sa počet obyvateľov odkázaných na nejaký druh dlhodobej starostlivosti sa bude dostávať do popredia záujmu. Ako sme uviedli, dlhodobá starostlivosť si bude vyžadovať koncepčné riešenie a integrované poskytovanie sociálnych a zdravotných služieb. Tento proces prebieha vo všetkých krajinách Európskej únie, pričom stav riešenia a integrácie systému dlhodobej starostlivosti v jednotlivých krajinách a na jednotlivých úrovniach je rôzna.

Literatúra

1. Vaňo, B, Bleha B.: Population projection of district in Slovakia until 2025, Infostat, Bratislava, 2008
2. Government decree No. 910/2000 on State health policy in the Slovak Republic
3. State policy of Health in the Slovak Republic, Government of the Slovak Republic, 2000.
4. Act No. 576/2004 on health care and services related to providing the healthcare
5. Act No. 578/2004 on providers of healthcare, medical workers and medical professional associations
6. Government ordinance No. 640/2008 on minimal public network of healthcare providers
7. Regulation No. 770/2004 of Ministry of Health, which determine characteristic sign of class of the individual medical facilities
8. Strategy of health care for geriatric patients and long-term patients, Government of the Slovak Republic, 2005.
9. Conception of social and long-term care in the Slovak Republic, Government of the Slovak Republic, 2005.
10. Regulation No. 364/2005 of Ministry of Health, which determine scale of nursing practice provided by nurse independently and with cooperation with physician and extent of birth assistance practice provided by birth assistant independently and in cooperation with physician
11. Act No. 447/2008 on financial allowances to compensation of several disabilities.
12. Act No. 448/2008 about social services
13. National priorities of development of social services, Ministry of Labour, Social Affairs and Family of the Slovak Republic, 2009.
14. The 2009 Ageing Report, European Commission

Kontakt

Ing. Marek Radvanský
Ekonomický ústav SAV
Šancova 56, Bratislava
Marek.Radvansky@savba.sk

Inteligentné technológie – genetický algoritmus Intelligent Technologies – Genetic algorithm

Tomáš Moravčík

Abstract: This article focuses on basic introduction to genetic algorithms. It explains main parts of the algorithm and also focuses more deeply to key features of the algorithm. Article is extended with examples for smoother understanding of the topic by reader.

Key words: Genetic algorithm, evolutionary algorithm, reproduction, crossover, mutation, fitness function, selection, termination conditions, artificial intelligence, optimization...

Kľúčové slová: Genetický algoritmus, evolučný algoritmus, reprodukcia, kríženie, mutácia, fitness funkcia, selekcia, podmienky ukončenia, umelá inteligencia, optimalizácia...

1. Úvod

Genetický algoritmus (GA) - Je to špecializovaná technika používaná pre hľadanie kvázi optimálnych riešení v optimalizačných úlohách. K jej charakteristickým vlastnostiam patrí využitie výpočtovej sily, inšpirovanie sa v evolučnej biológii, abstrahovanie od zložitosti modelu (prístup čiernej skrinky), na úkor istoty že nájdeme optimálne riešenie sa uspokojujeme s kvázi optimálnym riešením o ktorom však vieme že síce nemusí byť najlepšie ale je veľmi dobré. Motiváciou na zavedenie takejto výpočtovej metódy je podobnosť s genetickou interpretáciou a platnosťou pravidiel prirodzeného výberu v živých populáciách. Tak ako v živých populáciách genetické informácie čiastočne determinujú živý organizmus, tak pri použití genetických algoritmov súbor parametrov (nad ktorými sa rieši optimalizačná úloha) determinuje aktuálny výstup modelu. Základné kamene prirodzeného výberu teda kríženie a mutácia sú taktiež základnými kameňmi genetických algoritmov a zabezpečujú nájdenie kvázi optimálneho riešenia po dosiahnutí určitého počtu krokov algoritmu.

2. Inicializácia

Pred spustením algoritmu prebehne nastavenie optimalizovaných parametrov modelu najčastejšie náhodne (pseudonáhodne). takýmto spôsobom sa vytvorí populácia obsahujúca k členov. V prvom kroku algoritmu sa pomocou tzv. fitness funkcie určuje kvázi optimálnosť toho ktorého jedinca. Pri úlohe hľadania maxima sa teda jedinec ktorého hodnota fitness funkcie vykáže najvyššie číslo považuje za najlepšie riešenie v kroku 1 pri úlohe hľadania minima funkcie je to samozrejme jedinec s najnižšou hodnotou fitness funkcie. Nasleduje aplikácie kríženia a mutácie pomocou vopred určeného kľúča nad jednotlivými jedincami populácie.

Pre názornosť navrhujeme interpretovať jednotlivé kroky na príklade hľadania minima funkcie $f(x,y) = x^2 + y^2$. V tabuľke 1 sú uvedené náhodne inicializované hodnoty parametrov x, y , v tretom stĺpci je vyčíslená hodnota fitness funkcie.

Tabuľka 1: Inicializácia

	X	Y	Fitness $f(x,y) = x^2 + y^2$
Jedinec A (rodič A)	4	5	$16 + 25 = 41$
Jedinec B (rodič B)	2	7	$4 + 49 = 51$

3. Reprodukcia

Pri krížení sa vyberú najčastejšie dvaja jedinci (rodičia), z každého rodiča sa vyberú náhodne hodnoty premenných (génov), ktoré sa následne použijú pri tvorbe nového jedinca (dieťaťa), teda vytvárame nového člena populácie. Najčastejšie prichádza ku kríženiu pomocou rovnomerného štatistického rozdelenia a teda obaja rodičia sa podieľajú na potomkovi štatisticky rovnakým dielom. Použitie práve dvoch rodičov je inšpirované biológiou, niektoré výskumné práce navrhujú využiť viac ako dvoch jedincov pri tvorbe dieťaťa [2][3].

Tabuľka 2: Kríženie

	X	Y	Fitnes $f(x,y) = x^2 + y^2$
Dieťa 1	4	7	$16 + 49 = 65$
Dieťa 2	2	5	$4 + 25 = 29$

Dieťa 1 je vytvorené pomocou X génu z rodiča 1 a Y génu z rodiča 2. Dieťa 2 je vytvorené pomocou X génu z rodiča 2 a Y génu z rodiča 1. Pri vyhodnotení fitness funkcie možno sledovať zhoršenie (zvýšenie hodnoty fitness oproti rodičom) v prípade prvého dieťaťa a naopak zlepšenie v prípade druhého dieťaťa.

Pri mutácií sa vyberú ľubovoľné (aj všetky) hodnoty parametrov čiže gény jedinca populácie a tieto sa náhodne modifikujú tentoraz však s použitím Gaussovho štatistického rozdelenia. Cieľom je upraviť hodnotu parametra o relatívne malú hodnotu $\pm a$ a teda hodnota parametra (génu) v ďalšom kroku bude $H \pm a$. Príklad mutácie možno sledovať v tabuľke 2.

Tabuľka 2: Mutácia

	X ($\pm a$)	Y ($\pm a$)	Fitnes $f(x,y) = x^2 + y^2$
Dieťa 3	$4 (+1) = 5$	$5 (-2) = 3$	$25 + 9 = 34$
Dieťa 4	$2 (-1) = 1$	$7 (+1) = 8$	$1 + 64 = 65$

Gén X Dieťaťa 3 je modifikovaný o hodnotu +1, gén Y Dieťaťa 3 je modifikovaný o hodnotu -2. Gén X Dieťaťa 4 je modifikovaný o hodnotu -1, gén Y Dieťaťa 4 je modifikovaný o hodnotu +1. Pri vyhodnotení fitness funkcie možno sledovať zhoršenie (zvýšenie hodnoty fitness oproti rodičom) v prípade druhého dieťaťa a naopak zlepšenie v prípade prvého dieťaťa.

4. Selekcia

Podstatnou otázkou pre beh genetického algoritmu je kľúč výberu jedincov do funkcií kríženia mutácie. Často ide o stochastickú metódu výberu jedincov, napríklad takzvaný ruletový výber kde sa jedinci zoradia vzostupne podľa hodnoty fitness funkcie ktorú dosiahli, následne sa im pridelia čísla na základe úspešnosti od 1 po k , stochastický výber spočíva v generovaní náhodného čísla q patriaceho do intervalu $\langle 1, k \rangle$ a následného výberu jedinca označeného indexom m . Iný spôsob selekcie je takzvaný turnajový výber, kde sa náhodne vyberú z populácie dvaja alebo viacerí jedinci a z nich sa následne vyberie jediniec s najlepšou hodnotou fitness funkcie. Použitie turnajového a ruletového výberu umožňujú prežiť aj relatívne slabým jedincom ktorí však môžu mať niektoré gény veľmi kvalitné. Používanie takzvaného elitárskeho výberu, kde sa vyberá len niekoľko najlepších jedincov sa neodporúča práve kvôli zvýšenej pravdepodobnosti možnosti uviaznutia v lokálnom extréme.

Rozhodnutie medzi koľkými jedincami nastane kríženie, koľko jedincov bude mutovaných a koľko členov populácie bude nahradených novými jedincami (deťmi) v ďalšom kroku algoritmu (počet členov populácie ostáva v jednotlivých krokoch spravidla konštantný) ostávajú na tvorcovi algoritmu a ich modifikácia sleduje udržanie čo najvyššieho selektívneho tlaku a teda schopnosti algoritmu nájsť čo najlepšie riešenie za čo najmenší počet krokov a pritom neuviaznutie v lokálnom extréme. Modifikácia týchto parametrov algoritmu, počas jeho behu napr. na začiatku hľadania riešenia sa uprednostňuje použitie kríženia a pri konci hľadania sa preferuje použitie mutácie, prípadne variabilita pomeru detí a rodičov v rôznych krokoch algoritmu patria do tzv. genetických stratégií, kde už neoptimalizujeme len hodnoty parametrov modelu ale snažíme sa zvyšovať tzv. selektívny tlak aplikovaním genetického algoritmu aj na parametre samotného algoritmu.[1]

5. Ukončenie algoritmu

Kritérium ukončenia vyhľadávania môže byť niektoré z nasledujúcich, prípadne ich kombinácia:

- dosiahnutý počet nových generácií
- uspokojenie optimalizačného kritéria
- inšpekcia zlepšovania fitnes funkcie
- manuálna inšpekcia

6. Záver

Genetické algoritmy aj napriek neschopnosti poskytnúť optimálne riešenie v optimalizačných úlohách predstavujú robustnú alternatívu k tradičným optimalizačným technikám. Ich charakteristické vlastností ich predurčujú k využívaniu práve v prípadoch, kde analytické metódy zlyhávajú prípadne neposkytujú postačujúce výsledky. Ide predovšetkým o prípady kde sa optimalizuje veľké množstvo premenných alebo optimalizovaná funkcia má veľké množstvo lokálnych extrémov. Spolu s neurónovými sieťami a fuzzy systémami tvoria jeden z pilierov inteligentných systémov používaných v súčasnosti.

7. Literatúra

- [1]. EIBEN, A.E - and SMITH, J.E, Introduction to Evolutionary Computing Corr. 2nd printing, 2007, ISBN: 978-3-540-40184-1
- [2] EIBEN, A.E et al. Genetic algorithms with multi-parent recombination. PPSN III: Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature, 2005, 78-87. ISBN 3-540-58484-6.
- [3] TING, CHUAN-KANG. On the Mean Convergence Time of Multi-parent Genetic Algorithms Without Selection, 2005. Advances in Artificial Life: 403-412. ISBN 978-3-540-28848-0.

Adresa autora

Tomáš Moravčík Ing.
Slovenska Technická Univerzita
Fakulta elektrotechniky a informatiky
Íľkovičová 3, Bratislava
moravcik.tomas@gmail.com

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

Využitie logistickej regresie pri modelovaní chorobných stavov v gynekológii

Application of logistic regression for modelling of pathological phenomenon in gynaecology

Iveta Stankovičová, Maroš Suchánek, Jozef Chajdiak

Abstract: Binomial (binary) logistic regression is a form of regression, which is used when the dependent is a dichotomy and the independents are of any type. Logistic regression can be used to predict a dependent variable on the basis of continuous and/or categorical independents. We will apply logistic regression model for modelling of pathological phenomenon in gynaecology. In the concrete, we will apply logistic regression for modelling risk factors for indication abruptio placentae praecox.

Abstrakt: Binárna logistická regresia je druh regresného modelu, ktorý sa používa v prípadoch, kedy závislá premenná je binárna a nezávislé premenné sú rôznych typov. Logistická regresia predikuje hodnoty závislej premennej na základe kombinácie hodnôt vysvetľujúcich spojitých a/alebo kategoriálnych premenných. V článku budeme využívať model logistickej regresie na modelovanie chorobných stavov v gynekológii. Konkrétne pôjde o aplikáciu logistickej regresie pri modelovaní rizikových faktorov predčasného odlúčenia placenty (*lat. abruptio placentae praecox*).

Key words: gynaecology, abruptio placentae praecox, logistic regression, logit, odds ratio, risk factor, software SAS®

Kľúčové slová: gynekológia, predčasné odlúčenie placenty, logistická regresia, logit, pomer šancí, rizikový faktor, softvér SAS®

1. Úvod

Medzi jeden z najťažších stavov v pôrodníctve patrí predčasné odlučovanie placenty rodičky (*lat. abruptio placentae praecox*). Je to chorobný stav, ktorý patrí medzi najčastejšie príčiny úmrtia matiek a ich detí. U dieťaťa môže dochádzať k prerušeniu dodávky kyslíka a u matky môže dôjsť ku veľkým stratám krvi. Predčasné odlučovanie lôžka je vždy neočakávaná komplikácia a môže postretnúť ktorúkoľvek ženu. Táto porucha sa vyskytuje u 1 až 3 percent rodičiek v našej populácii.

V predkladanom príspevku sa pokúsime aplikovať model logistickej regresie na dáta z oblasti gynekologického výskumu, v ktorom bol sledovaný jav predčasného odlúčenia placenty rodičky a množstvo sprievodných javov a symptómov (resp. premenných, faktorov). Budeme modelovať výskyt odrhnutia resp. odlúčenia lôžka maternice v priebehu gravidity v závislosti od rôznych rizikových faktorov. Naším cieľom je ukázať, že logistická regresia má efektívne využitie v medicínskom výskume a pri správnej aplikácii môže priniesť zaujímavé a užitočné výsledky. Na výpočty sme použili softvér SAS®.

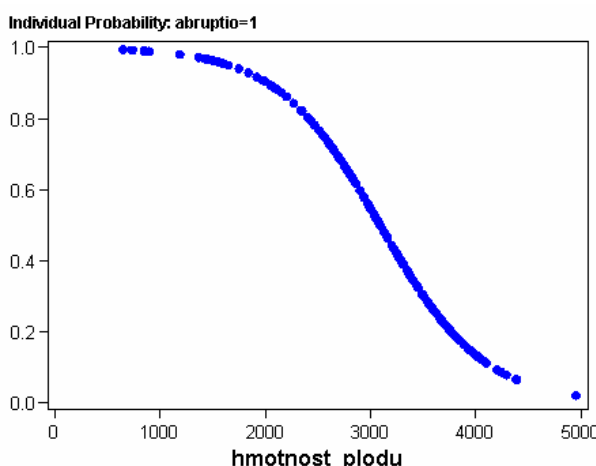
2. Model binárnej logistickej regresie

Logistická regresia je veľmi podobná lineárnej regresii. Aj logistická regresia je založená na štatistickom rozdelení a patrí medzi robustné nástroje pre tvorbu modelov. Hlavný rozdiel je v tom, že závislá premenná nie je spojitá, ale je diskretná, resp. kategoriálna. Aby sa dala použiť regresia, tak sa závislá premenná transformuje na spojitú

hodnotu, ktorá je funkciou pravdepodobnosti výskytu udalosti. Logistická regresia je špeciálnym prípadom zovšeobecného lineárneho modelu (*GLM = General Linear Model*).

Označme ako $\mathbf{x}$ vektor nezávislých premenných. Označme ako Y binárnu závislú premennú, ktorá nadobúda dve možné hodnoty 0 (neúspech, resp. želaná udalosť nenastala) a 1 (úspech, resp. želaná udalosť nastala). Označme symbolom p pravdepodobnosť výskytu želanej udalosti $Y=1$, teda hodnotu, ktorú modelujeme. Potom $p = P(Y=1/\mathbf{x})$ je podmienená pravdepodobnosť výskytu želanej udalosti za podmienky výskytu vektora nezávislých premenných $\mathbf{x}$. Podiel $p/(1-p)$ vyjadruje šance (odds) výskytu želanej udalosti (úspechu, $Y=1$) k výskytu neželanej udalosti (neúspechu, $Y=0$). Šance a pravdepodobnosť predstavujú tie isté informácie, len v iných podobách, resp. inými slovami.

Vzťah medzi pravdepodobnosťou p a vysvetľujúcou premennou X nie je lineárny. Grafickým zobrazením závislosti je logistická sigmoidná krivka, tzv. S-krivka (*Obrázok 1*).



Obrázok 1: S-krivka vyjadrujúca závislosť pravdepodobnosti odlúčenia lôžka počas tehotnosti (*abruptio=1*) a hmotnosti plodu

Pre finálnu transformáciu závislej premennej na spojitú premennú sa vypočíta logaritmus šancí, tzv. **logit**(p) = $\ln(p/(1-p))$. Vzťah medzi logitom a vektorom vysvetľujúcich premenných má už lineárny charakter. Rovnica logistického modelu má nasledovný tvar:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (1)$$

Vzťah medzi pravdepodobnosťou a vektorom vysvetľujúcich premenných dostaneme spätnou transformáciou a má nelineárny charakter. Je to druh exponenciálnej funkcie:

$$p = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}} \quad (2)$$

Logitovou transformáciou dostávame z nelineárnej závislosti závislosť lineárnu. Kým hodnoty pravdepodobností p_i sú z intervalu (0,1), hodnoty podielov $p/(1-p)$ sú nezáporné hodnoty, tak hodnoty logitov môžu nadobúdať akékoľvek reálne hodnoty z intervalu $(-\infty, \infty)$.

Okrem najčastejšie používanej logitovej transformácie sa v teórii vyskytujú aj iné druhy transformácií, tzv. linkových funkcií, v logistických modeloch. Systém SAS ponúka okrem logitu aj probit (resp. normit) a komplementárnu log-log funkciu. Prehľad vzorcov pre tieto transformácie podmienenej pravdepodobnosti p sú nasledovné:

- Logitová funkcia má tvar $g(p) = \ln(p/(1-p))$ a k nej inverzná funkcia, čiže jej distribučná funkcia, je $F(x) = 1/(1+\exp(-x)) = \exp(x)/(1+\exp(x))$.

- Probitová (resp. aj normit) funkcia má tvar $g(p) = \Phi(x)^{-1}$. $\Phi(x)$ v tomto prípade znamená distribučnú funkciu normovaného normálneho rozdelenia. Probitová funkcia zvyčajne obsahuje aj pripočítanú konštantu 5.

- Komplementárna log-log funkcia má tvar $g(p) = \ln(-\ln(1-p))$. Jej inverziou sa získa distribučná funkcia, ktorá sa tiež volá Gompertzova funkcia a rovnica má nasledovný tvar: $F(x) = 1 - \exp(-\exp(x))$.

3. Odhad parametrov modelu logistickej regresie

Modely logistickej regresie môžu obsahovať rôzne druhy vysvetľujúcich premenných (prediktorov X_i , resp. rizikových faktorov). Charakter vysvetľujúcich premenných je podstatný pre konštrukciu modelu, pre odhad jeho parametrov, hodnotenie jeho kvality a aj využitie. Dôležité je, či sú kombinácie vysvetľujúcich premenných jedinečné (tak ako je to pri spojitých premenných), a tak môžeme na konštrukciu logistického modelu použiť netriedené údaje. Keď uvažujeme len s vysvetľujúcimi premennými kategoriálneho typu, tak je možné vytriediť údaje do viacrozmernej kontingenčnej tabuľky a potom ich použiť pre modelovanie. Logistická regresia sa v prípade, že nezávislou premennou je len jedna kategoriálna premenná, redukuje na analýzu kontingenčnej tabuľky.

Na odhad parametrov modelu logistickej regresie sa dnes najčastejšie používa metóda maximálnej vierohodnosti. Táto metóda sa mohla začať používať až s rozvojom výpočtovej techniky, lebo je výpočtovo veľmi náročná. Je vhodná na odhad parametrov nelineárnych modelov, medzi ktoré patrí aj logistická regresia. V minulosti sa na odhad parametrov logistických modelov používala aj metóda diskriminačnej funkčnej analýzy, ktorá vyžaduje podmienku normality rozdelenia pre vysvetľujúce premenné (tak ako metóda najmenších štvorcov). V prípade, že niektorá vysvetľujúca premenná je kategoriálna, metóda diskriminačnej funkčnej analýzy dáva skreslené (resp. vychýlené) výsledky, zvyčajne sa to prejaví tak, že odhady pomerov šancí (*odds ratio*) sú veľmi vysoké. Metóda maximálnej vierohodnosti (ML) nemá žiadne požiadavky na nezávislé premenné a tie môžu byť nominálne, poradové (ordinálne), alebo intervalové (diskrétne alebo spojité).

4. Stratégia tvorby modelu logistickej regresie

Dobrý a kvalitný model logistickej regresie vytvoríme len vtedy, keď dodržiavame nasledovný postup (stratégiu):

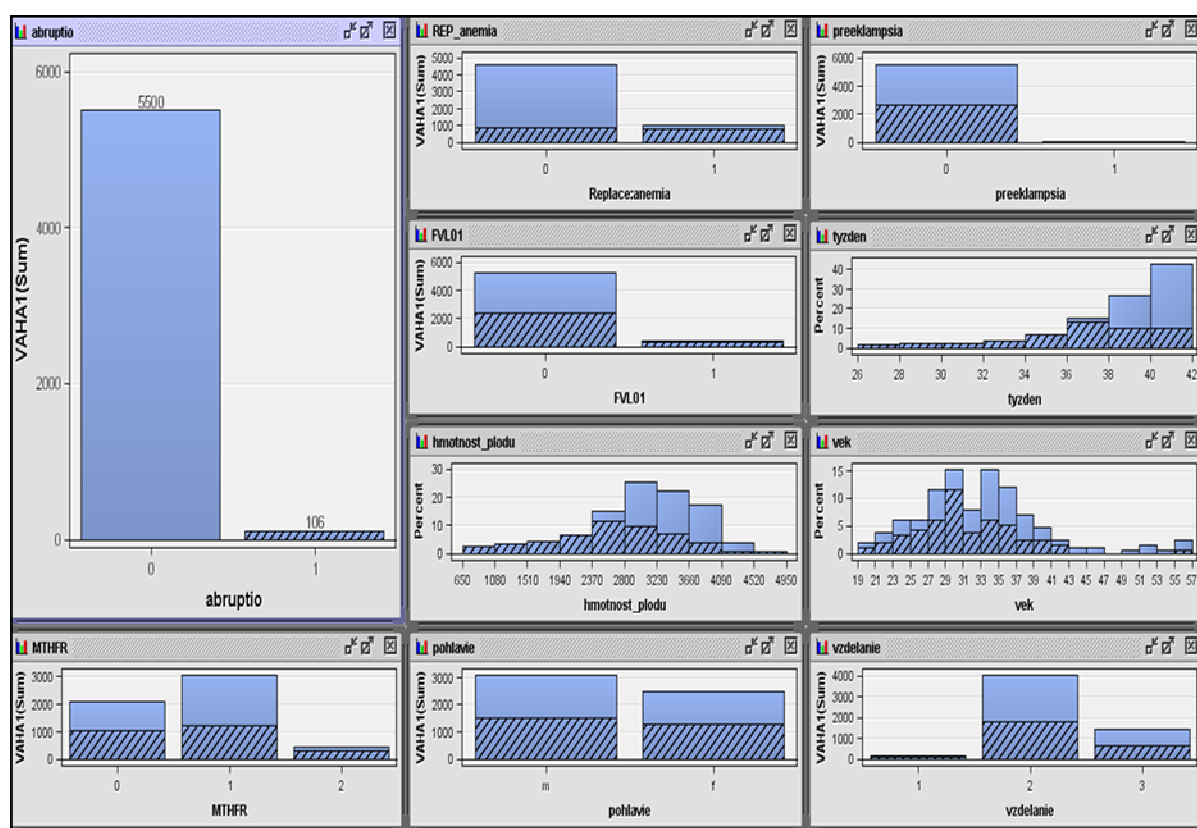
1. Urobíme jednorozmernú analýzu všetkých premenných.
2. Nakreslíme si grafy logitov pre spojité a poradové vysvetľujúce premenné a analyzujeme ich.
3. Vykonáme stratifikovanú analýzu na identifikáciu interakcií medzi vysvetľujúcimi premennými a na identifikáciu „mäťúcich“ (*confounders*) premenných.
4. Použijeme rôzne selekčné metódy na identifikáciu niekoľkých prijateľných modelov.
5. Identifikujeme extrémne (*outliers*) a vplyvné (*influential*) hodnoty.
6. Zhodnotíme kvalitu výsledného modelu, resp. modelov.

5. Modelovanie chorobných stavov pri pôrode pomocou logistickej regresie

Súbor analyzovaných údajov z oblasti gynekológie má celkový rozsah 216 pozorovaní ($n=216$) a skladá sa z dvoch častí. V prvej časti ($n_1=106$) sú údaje o ženách-rodičkách, u ktorých sa vyskytla sledovaná porucha, čiže predčasné odlúčenie lôžka matrice (premenná *abruptio*=1). Druhú časť ($n_2=110$) predstavujú údaje o ženách-rodičkách z kontrolnej skupiny (premenná *abruptio*=0). Binárna premenná *abruptio* vystupuje v modeli logistickej regresie ako závislá čiže modelovaná premenná a budeme modelovať výskyt sledovanej poruchy, čiže jej hodnotu 1 (*abruptio*=1).

Ako rizikové faktory pre predčasné odlúčenie lôžka maternice u žien – rodičiek boli sledované mnohé rizikové faktory. Spolu ich bolo v poskytnutom súbore údajov až 45. Tieto faktory môžeme rozdeliť do niekoľkých skupín. Do prvej skupiny patria faktory, ktoré charakterizujú zdravotný stav ženy – rodičky, napr. jej vek a hmotnosť na začiatku a na konci gravidity, prítomnosť rôznych chorôb a porúch (anémia, cukrovka, vysoký krvný tlak, dedičné mutácie génov, apod.), ale aj jej sociálne postavenie a správanie: vzdelanie, fajčenie, alkohol a podobne. Do druhej skupiny môžeme zaradiť vplyvy, ktoré mohli alebo môžu poškodzovať sliznicu dutiny maternice, napr. predchádzajúci pôrod, cisársky rez, spontánny alebo umelý potrat a podobne. Do tretej skupiny patria faktory poškodzujúce kvalitu cievnej steny, ako fajčenie, cukrovka, vysoký krvný tlak a podobne. Do samostatnej skupiny patria charakteristiky plodu, napr. jeho hmotnosť a dĺžka, týždeň pôrodu, pohlavie a podobne.

V analýze sme použili aj nami vytvorenú frekvenčnú premennú *vaha* (*vaha*=1 pre súbor údajov kde *abruptio*=1, *vaha*=50 pre súbor údajov kde *abruptio*=0), aby sme dosiahli „reálny“ výskyt sledovanej poruchy, a to približne 2 percentá z populácie (viď Obrázok 2, $106 : 5500 = 2\%$).



Obrázok 2: Výskyt poruchy (*abruptio*=1) podľa vybraných rizikových faktorov

V procese modelovania sme použili najprv vecnú analýzu a potom štatistickú analýzu (metóda selekcie premenných *Stepwise*) na identifikáciu, resp. výber rizikových faktorov pri výskyte sledovanej poruchy. Pri identifikácii nám pomohla aj jednorozmerná grafická analýza histogramov a stĺpcových grafov (viď Obrázok 2). Je zrejmé, že u žien – rodičiek s anémiou sa skoro vždy vyskytuje aj odlúčenie maternice. Podobná situácia nastáva aj v prípadoch, keď rodičky majú tehotnosťou vyvolaný vysoký krvný tlak a prítomnú genetickú mutáciu faktora V Leiden. Tiež môžeme sledovať, že s vyššou hmotnosťou plodu, s vyšším týždňom tehotnosti a s rastúcim vekom rodičiek výskyt sledovanej poruchy klesá.

Ako štatisticky významné vysvetľujúce premenné sme nakoniec identifikovali tieto rizikové faktory (*Poznámka*: Faktory sú zoradené podľa ich štatistickej významnosti, viď *Tabuľka 1*):

- *REP_anemia* – chudokrvnosť rodičky: 1 - prítomná, 0 – neprítomná;
- *tyzden* - týždeň tehotnosti, kedy nastalo odtrhnutie lôžka resp. pôrod v skupine zdravých rodičiek;
- *hmotnost_plodu* – pôrodná hmotnosť plodu (g) v čase odtrhnutia lôžka (u žien so sledovanou poruchou automaticky pôrod cisárskym rezom), alebo hmotnosť plodu pri pôrode v referenčnej skupine zdravých žien;
- *MTHFR* – mutácia génu enzýmu metyléntetrahydrofolátreduktázy, ktorá zvyšuje krvnú zrážanlivosť: 0 - mutácia neprítomná, 1 - heterozygotná mutácia (1 gén zdravý a 1 chorobný), 2 - homozygotná mutácia (oba chorobné gény, od oboch rodičov);
- *pohlavie* - pohlavie plodu: f - ženské, m – mužské;
- *vzdelanie* - vzdelanie rodičky: 1 - základné, 2 - stredné, 3 – vysokoškolské;
- *preeklampsia* - tehotnosťou vyvolaný vysoký krvný tlak;
- *FVL01* - faktor V Leiden, ktorý znamená dedičnú mutáciu 5. faktora krvnej zrážanlivosti a zvyšuje tendenciu ku krvnej zrážanlivosti, možnosti: 0 – mutácia neprítomná, 1 - mutácia prítomná (heterozygotná alebo homozygotná mutácia prítomná);
- *vek* – vek rodičky v rokoch.

Tabuľka 1: Významnosť premenných v modeli

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
REP_anemia	1	86.77	<.0001
tyzden	1	34.27	<.0001
hmotnost_plodu	1	22.28	<.0001
MTHFR	2	19.96	<.0001
pohlavie	1	15.37	<.0001
vzdelanie	2	11.92	0.0026
preeklampsia	1	8.07	0.0045
FVL01	1	7.46	0.0063
vek	1	5.59	0.0181

Tabuľka 2: Odhady pomerov šancí

Odds Ratio Estimate	
Effect	Point Estimate
FVL01 0 vs 1	0.265
MTHFR 0 vs 2	0.880
MTHFR 1 vs 2	0.160
REP_anemia 0 vs 1	0.022
hmotnost_plodu	0.998
pohlavie f vs m	4.417
preeklampsia 0 vs 1	0.107
tyzden	0.440
vek	0.947
vzdelanie 1 vs 3	1.848
vzdelanie 2 vs 3	0.435

Výsledný logistický model je možné dobre interpretovať pomocou pomerov šancí (stĺpec *Odds Ratio Point Estimate*, *Tabuľka 2*). Ako najvýznamnejší rizikový faktor nám vyšla prítomnosť anémie. V prípadoch, keď je rodička anemická (*anemia=1*), tak jej hrozí odtrhnutie lôžka s priemernou šancou vyššou ako 45:1 (lebo $1/0.022=45.45$) v porovnaní s rodičkou, ktorá nemá anémiu (*anemia=0*). Na druhom mieste v našom modeli sa nachádza rizikový faktor týždeň gravidity. Z výsledného pomeru šance (0.44) je zrejmé, že s každým ďalším týždňom gravidity priemerná šanca, že nastane sledovaná porucha, klesá v pomere 0.44:1. Podobne sa správajú aj rizikové faktory vek rodičky (priemerne 0.947:1 na každý prírastok veku o 1 rok) a hmotnosť plodu (priemerne 0.998:1 na každý prírastok hmotnosti plodu o gram).

V prípade výskytu genetických porúch u rodičiek sme identifikovali ako významné dve poruchy. Ak sa u rodičky vyskytuje mutácia oboch génov enzýmu metyléntetrahydrofolátreduktázy, ktorá zvyšuje krvnú zrážanlivosť (*MTHFR=2*), tak šance na

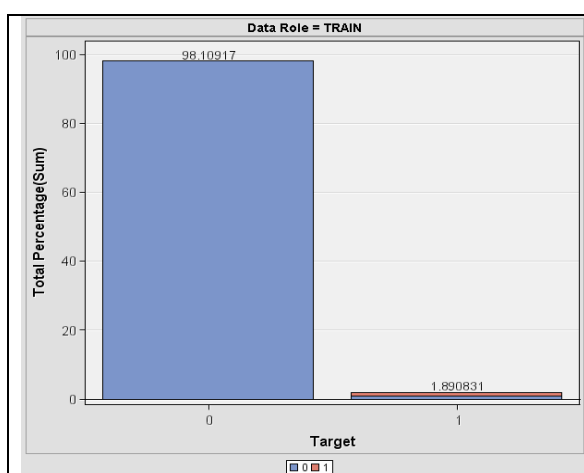
výskyt predčasného odlúčeného lôžka maternice sú vyššie, konkrétne v porovnaní len s mutáciou 1 génu sú šance vyššie v priemere až 6.25:1 ($1/0.016=6.25$). Zaujímavé je, že keď porovnávame stav mutácie oboch génov ($MTHFR = 2$) so stavom, keď pacientka nemá prítomnú túto mutáciu ($MTHFR = 0$), tak je pomer šancí výrazne nižší, len 1.14:1 ($1/0.88=1.14$). Tento výsledok, podľa nášho názoru, vyžaduje ďalšie skúmanie, ktoré by to potvrdilo alebo vyvrátilo. V prípade výskytu dedičnej mutácie 5. faktora krvnej zrážanlivosti ($FVL01=1$) u rodičky, je priemerná šanca na odlúčenie lôžka 3.77:1 ($1/0.265=3.77$). Aj graviditou vyvolaný vysoký krvný tlak je významný rizikový faktor sledovanej poruchy ($preeklampsia=1$). U rodičky s týmto chorobným symptómom je šanca na odlúčenie lôžka až 9.35-krát vyššia ($1/0.107=9.35$) v porovnaní s rodičkou, ktorá nemá tento symptóm ($preeklampsia=0$).

Zaujímavé je, že pohlavie plodu, je tiež rizikový faktor pre sledovanú poruchu. Ak plod je ženského pohlavia ($pohlavie=f$), tak rodička má vyššiu šancu na odlúčenie lôžka, a to až 4.4:1 v porovnaní s tým, keď plod je mužského pohlavia ($pohlavie=m$). Aj vzdelanie rodičky sa ukázalo ako významný rizikový faktor. Z výsledkov je zrejmé, že poruchou predčasného odlúčeného lôžka trpia hlavne ženy so základným vzdelaním.

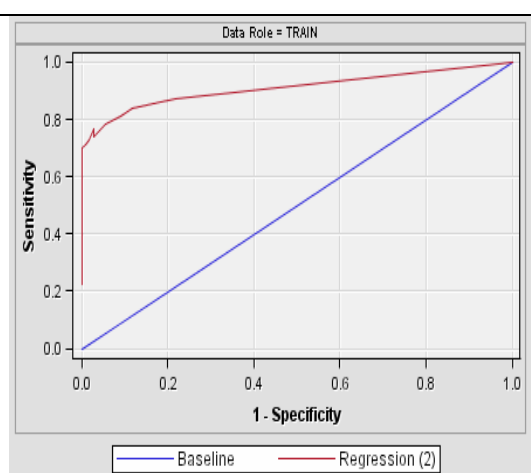
Čo sa týka kvality prezentovaného modelu logistickej regresie, je potrebné povedať, že model dobre predikuje prípady, kedy sledovaná porucha určite nenastane ($abruptio=0$). Avšak pri predikcii výskytu sledovanej poruchy ($abruptio=1$) je model úspešný len viac ako v polovici prípadov (55.66%). Z celkového počtu 106 prípadov, kedy došlo k odlúčeniu lôžka vo vzorke ($abruptio=1$), model predikoval len 59 prípadov ako $abruptio=1$ a až 47 prípadov predikoval ako $abruptio=0$ (viď Tabuľka 3, Obrázok 3). Aj z grafu ROC-krivky je zrejmé, že výsledný model nie je ideálny. I keď ROC index má vysokú hodnotu ($c=0.93$), ale priebeh ROC krivky nie je plynulý.

Tabuľka 3: Kvalita modelu logistickej regresie

Target	Target Outcome	Outcome Percentage	Frequency Percentage	Total Count	Percentage
0	0	99.15	100.00	5500	98.11
1	0	0.85	44.34	47	0.84
1	1	100.00	55.66	59	1.05



Obrázok 3: Podiel chybných klasifikácií modelu



Obrázok 4: ROC krivka

6. Záver

Modely logistickej regresie majú v praxi, a aj v medicínskej praxi, široké použitie. Pri ich vytváraní treba dodržať určité postupy. V úvode je potrebná podrobná vecná analýza problému a potom nadväzuje štatistická analýza. Dôležitý je vecný výber vysvetľujúcich premenných do modelu. Na začiatku je potrebné posúdiť, ktoré sledované premenné sú príčinou, a ktoré dôsledkom modelovaného javu. Pri tom nám pomáha aj jednorozmerná exploračná a grafická analýza. Dôležitá je aj vhodná kategorizácia premenných, tých ktoré sa budú používať ako vysvetľujúce premenné (faktory), lebo od toho mnoho ráz závisí ich predikčná schopnosť a využiteľnosť v modeloch.

Výsledný logistický model z oblasti gynekológie, ktorý modeluje rizikové faktory predčasného odtrhnutia lôžka maternice, odhalil niektoré rizikové faktory a kvantifikoval ich vplyv na modelovanú premennú. Výsledný model však nemôžeme považovať za konečný, lebo kvalita predpovede modelovanej hodnoty *abruptio=1* je len o niečo vyššia ako 50%. Tento fakt mohol byť spôsobený napríklad pomerne malou vzorkou dát, v ktorej sa niektoré kombinácie rizikových faktorov s modelovanou hodnotou závislej premennej *abruptio=1* vyskytovali s nízkou frekvenciou. Odporúčame pokračovať v zbieraní údajov o sledovanej poruche a potom znovu model prepočítať.

7. Literatúra

- [1] ALBERT, A., ANDERSON, J.A. : *On the Existence of Maximum Likelihood Estimates in Logistic Regression Models*. In: Biometrika, 71/1984, p. 1-10.
- [2] CHAJDIK J.: *Štatistické porovnanie zhody analyzovaného a referenčného súboru*. In: Forum Statisticum Slovaca 6/2008, str. 63-66. SŠDS Bratislava 2008. ISSN 1336-7420.
- [3] KLEINBAUM, DAVID G.: *Logistic regression*. A Self-Learning Text. Springer. 1994
- [4] LABUDOVÁ V.: *Využitie prediktívnych modelov v medicínskej praxi*. In: 13. ročník konferencie SAS Forum 2010. Praha. Dostupné na internete (18.3.2010): <http://www.sas.com/reg/offer/cz/2010_sas_forum_2010?page=vs>
- [5] LUHA, J.: *Štatistické metódy analýzy kvalitatívnych znakov*. In: Forum Statisticum Slovaca 2/2006. SŠDS Bratislava 2006. ISSN 1336-7420.
- [6] ŘEZANKOVÁ, H.: *Analýza dat z dotazníkových šetření*. SPSS Profesional Publishing. Praha. 2007. ISBN 978-80-86946-49-8.
- [7] ŘEZANKOVÁ, H.: *Analýza kategoriálních dat*. Nakladatelství Oeconomica. VŠE Praha. 2005. ISBN 80-245-0926-01
- [8] SHARMA, S.: *Applied Multivariate Techniques*. New York, John Wiley & Sons, Inc., 1996.
- [9] STANKOVIČOVÁ, I., VOJTKOVÁ, M.: *Viacrozmerné štatistické metódy s aplikáciami*. Iura Edition. Bratislava, 2007.
- [10] STANKOVIČOVÁ, I.: *Grafy interakcií. Zobrazovanie výsledkov logistických modelov*. In: Výpočtová štatistika. Bratislava : SŠDS, 2003, s. 105-109. ISBN 80-88946-29-8
- [11] STANKOVIČOVÁ, I.: *Logická regresia a hĺbková analýza dát (Data Mining)*. In: Štatistické metódy v praxi. Bratislava: SŠDS 2002. ISBN 80-88946-19-0
- [12] STANKOVIČOVÁ, I.: *Logistická regresia a jej využitie v ekonomickej praxi*. In: Forum Statisticum Slovaca. - Roč. 3, č. 1/2007, s. 42-54. ISSN 1336-7420.

Adresy autorov:

Iveta Stankovičová, Ing., PhD.
Katedra informačných systémov
Fakulta manažmentu UK
Bratislava
iveta.stankovicova@fm.uniba.sk

Maroš Suchánek, MUDr.
FN Ružinov
Bratislava
marossuchanek@gmail.com

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

Výpočet normatívnych hodnôt intima-medie na artérii carotis communis

Calculation normative values of intima-medie on carotis communis artery

Madarász Štefan, Chajdiak Jozef

Abstract: Demographic readings highlight on the fact, that in the Central European's states, including Slovak republic, morbidity and mortality on cardiovascular (KVO) and cerebrovascular diseases show almost triple higher values than in the developed countries of West Europe. The most common reason of these diseases is atherosclerosis, which has a long-term process of development, and we can detect its individual stages by measure a thickness of intima medie carotids by means of ultrasonograph. On appraisal of finding is needed to have a scale of normal values in individual age groups. In this work the authors highlight on the possibility of calculate a range of normal values by means of programmatic bundle MS Office, Excel, subsystem PivotTable.

Key words: Cardiovascular diseases, Cerebrovascular diseases, Intima-Media Artery Carotis communis, normal values CIMT, MS-Office, Excel- Subsystem PIVOTTABLE

Kľúčové slová: kardiovaskulárne ochorenia (KVO) - cerebrovaskulárne ochorenia(CVO) - intima-media arterie carotis communis (CIMT) - normálne hodnoty CIMT-MS-Office, Excel - Subsystem PivotTable.

1. Úvod

Štatistické údaje poukazujú na alarmujúcu skutočnosť. Od začiatku minulého storočia výskyt chorôb obehového systému vykazuje narastajúcu tendenciu, a podľa Svetovej zdravotníckej organizácie (WHO) kardiovaskulárne (KVO) a cerebrovaskulárne (CVO) ochorenia zapríčiňujú takmer 30% úmrtia obyvateľstva v celosvetovom meradle. (Madarász, 2008).

Najčastejšou príčinou týchto ochorení je ateroskleróza. Ateroskleróza je proces dlhodobý, mnoho krát začína už v 2 dekáde, a počas rokov svojho vývoja dospeje k vyvolaniu svojich závažných komplikácii ktoré sa prejavujú ťažkými postihnutiami tepenného systému cirkulácie mnoho krát až s fatálnym koncom.

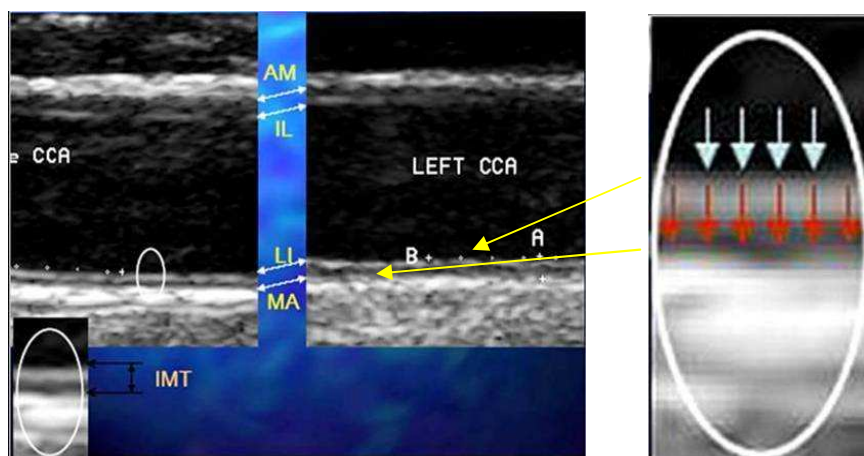
Žiaľ, táto situácia neobišla ani Slovenskú republiku, kde počet úmrtí na CHOS ďaleko prevyšuje európsky priemer. (Madarász, 2008). V Európe umiera 1,9 miliónov osôb ročne a z toho 42 % na kardiovaskulárne choroby. Náklady na KVO činia ročne 169 miliárd eur, z toho ide 105 miliárd eur na liečbu a 64 miliárd eur stojí strata produktivity a následná starostlivosť. (Šimon, 2008).

Tieto skutočnosti podmienili potrebu hľadať spôsoby, ako včas zistiť objektívne rozvoj aterosklerotického procesu. (Bots, 2003)

Jednou z týchto metód je vyšetrenie intima-medie aretrie carotis communis (CIMT). (DeGrott, 2008, Linhart, 2003, Madarász, 2008, Spáčil, 2002).

Prvý krát pomocou ultrasonografickej technológií malú hrúbku intima-medie sa podarilo zobrazit' Pignolimu a spolupracovníkom. Intima-mediálnu šírku identifikovali ako

šírku echogenného rozhrania medzi lúmenom a intimou a echogenného rozhrania medzi médiou a adventíciou cievnej steny (DeGrott, 2008, Pignoli, 1986, Školoudík, 2005).



Obrázok 1: Zobrazená je prvá echogenná línia(A) a druhá echogenná línia (B) ktoré určujú IMT na vzdialenej strane cievnej steny od ultrazvukovej sondy (Csányi, 2005)

2. Cieľ práce

Stanovenie normálnych hodnôt intima-médie v populácii zdravých mužov a žien (klientov) vo vekových skupinách 30-40 r.; 41-50 r.; 51-60 r.; 61-70 r. a vo vekovej skupine viac ako 70 ročných mužov a žien na pravej a ľavej A. Carotis Communis (CIMT).

3. Materiál a metóda

Pacienti boli zaradení do vekových skupín 30-40 rokov, 41-50 rokov, 51-60 rokov, 61-70 rokov a viac ako 70 rokov. V každej vekovej kategórii bolo zaradených po 15 mužov a 15 žien, okrem vekovej skupiny viac ako 70 ročných pacientov, kde bolo zaradených len 12 mužov a 12 žien. Pacienti (klienti) boli bez akýchkoľvek subjektívnych ťažkostí, a základné odbery krvi (KO, diff., biochemický screening, moč) nevykazovali patologické hodnoty, doteraz neprekonali nejakú chorobu obehovej sústavy (CHOS), neboli u nich preukázané cirkulačné poruchy a neužívali nejaké lieky.

CIMT sa vyšetroval na USG prístroji Aloka 4000. Používala sa 12 Mhz lineárna sonda s vysokým rozlíšením a pri vyšetrení sa dodržali odporúčania Mannheim Intima Media Thickness Consensus (MIMTC). (Touboul 2004)

Z archivovanej snímky IMT v jpeg formáte sa semiautomatickou metódou pomocou softwaru Esaote IMT LAB 1.1 (merali IMT v oblasti ACC l.dx. a l.sin. u každého respondenta s presnosťou na tisícinu mm. (IMT lab, 2005)

Celkovo počet pacientov vo vyšetrovanej skupine bol 144. U každého pacienta sa vyšetrili tri krát pravá a ľavá arteria karotis communis. Celkový počet vyšetrení CIMT bol 864, z toho počet vyšetrení ľavej karotídy bol 432. Celkove sa vyšetrilo 72 žien a 72 mužov.

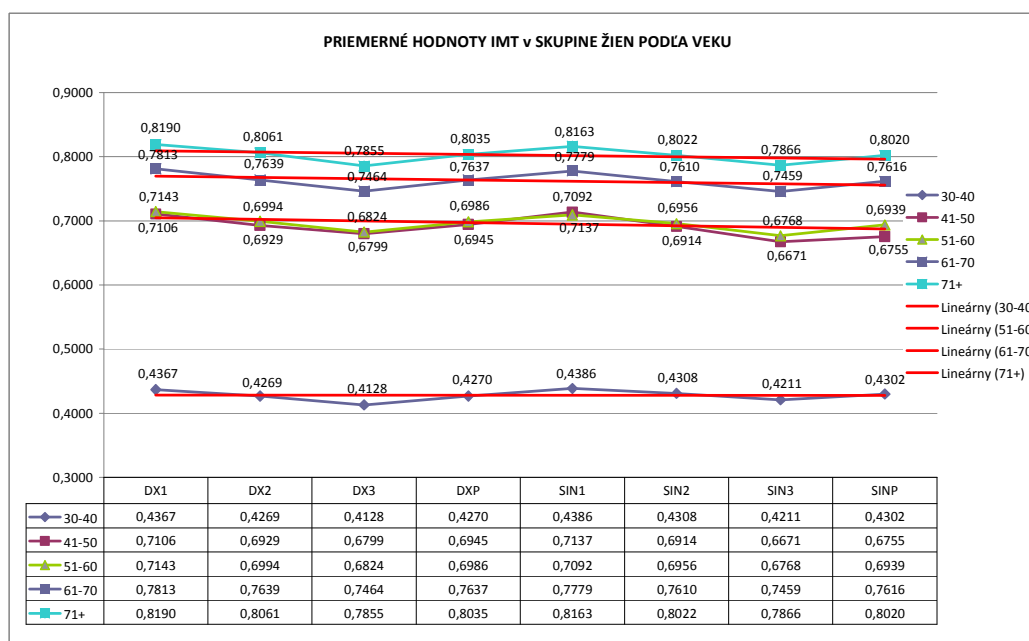
Výsledky jednotlivých meraní boli zaznamenané do tabuliek, kde sa vypočítali priemerné hodnoty v jednotlivých skupinách pacientov z troch meraní IMT na pravej a na ľavej arterii carotis communis (ACC). V každej vekovej skupine sa hodnotila priemerná hrúbka intima-médie v celom súbore pacientov, a samostatne v skupine mužov a u žien. Taktiež sa samostatne hodnotili priemerné hodnoty na pravej a na ľavej strane a. carotis communis. Štatistické hodnotenie bolo vykonané pomocou MS Office. Vypočítala sa

priemerná hrúbka CIMT z troch meraní na pravej a na ľavej acc u mužov a u žien. Pomocou softvéru MS Office, Excel, funkcia STDEV sa vypočítali smerodajné odchýlky a slúžili na posúdenie miery variability a pomocou MS Office, excel, podsystémom PivotTable sa hodnotili percentily.

4. Výsledky

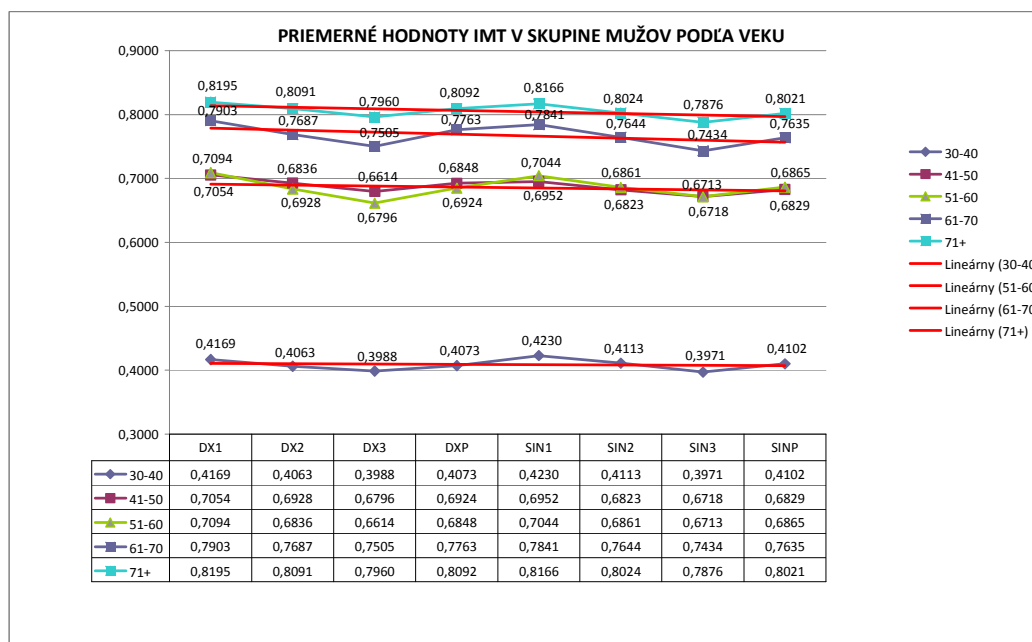
Na grafe č. 2. je vidieť, že narastanie hrúbky intima-medie vykazuje významnú závislosť od veku pacienta. Vo vekovej skupine 30-40 ročných pacientiek dosahuje najmenšie hodnoty. Vo vekovej skupine 41-50 a 51-60 ročných žien je prítomný nárast hrúbky intima-medie, ale medzi týmito dvomi vekovými skupinami nie je prítomný významnejší rozdiel.

Podobne ako na predchádzajúcom grafe, aj na grafe č. 30 je vidieť, že narastanie hrúbky intima-medie vykazuje významnú závislosť od veku pacienta. Vo vekovej skupine 30-40 ročných mužov dosahuje najmenšie hodnoty. Vo vekovej skupine 41-50 a 51-60 ročných mužov je prítomný nárast hrúbky intima-medie, ale medzi týmito dvomi vekovými skupinami nie je prítomný významnejší rozdiel. Vo vekovej skupine 61-70 ročných pacientov a pacientov s viac ako 70 rokov je prítomný nárast hrúbky intima-medie v oboch skupinách. Aj tu je prítomná korelácia s literárnymi údajmi o vekovej závislosti zmeny hrúbky intima-medie. (17, 44, 69, 98, 114, 116)



Obrázok 2: Priemerné hodnoty IMT z troch meraní u žien podľa veku

Vo vekovej skupine 61-70 pacientov a pacientov s viac ako 70 rokov je prítomný nárast hrúbky intima-medie v oboch skupinách. Je to v súlade s literárnymi údajmi o vekovej závislosti zmeny hrúbky intima-medie. (Csiba, 2005, Školoudík 2003, 2006 Mannheim Intima-Media Thickness consensus).



Obrázok 3: Priemerné hodnoty IMT z troch meraní u mužov podľa veku

Podobne ako na predchádzajúcom grafe, aj na grafe č. 1 je vidieť, že nárast hrúbky intima-médie vykazuje významnú závislosť od veku pacienta aj u mužov. Vo vekovej skupine 30-40 ročných mužov dosahuje najmenšie hodnoty. Vo vekovej skupine 41-50 a 51-60 ročných mužov je prítomný nárast hrúbky intima-médie, ale medzi týmito dvomi vekovými skupinami nie je prítomný významnejší rozdiel. Vo vekovej skupine 61-70 ročných pacientov a pacientov s viac ako 70 rokov je prítomný nárast hrúbky intima-médie v oboch skupinách. Aj tu je prítomná korelácia s literárnymi údajmi o vekovej závislosti zmeny hrúbky intima-médie. (Csiba, 2005, Školoudík, 2004, Spáčil, 2002).

Smerodajné odchýlky boli vypočítané pomocou programu MS Office, Excel, funkcia STDEV a slúžili na posúdenie miery variability. (Chajdiak, 2003, Chajdiak, 2005)

Tabuľka 1: Smerodajné odchýlky u žien v jednotlivých vekových skupinách pri jednotlivých meraniach na pravej a na ľavej ACC

	ŽENY					F Total
Values	30-40 r.	41-50 r.	51-60 r.	61-70 r.	71+ r.	
StdDev of l.dx.I.	0,119	0,040	0,025	0,020	0,008	0,149
StdDev of l.dx.II.	0,115	0,040	0,027	0,019	0,011	0,147
StdDev of l.dx.III.	0,109	0,053	0,028	0,021	0,009	0,145
StdDev of l.dx.priemer	0,114	0,043	0,026	0,019	0,008	0,146
StdDev of l.sin.I.	0,106	0,047	0,021	0,017	0,008	0,146
StdDev of l.sin.II.	0,104	0,042	0,024	0,018	0,008	0,142
StdDev of l.sin.III.	0,102	0,049	0,032	0,022	0,007	0,141
StdDev of l.sin.priemer	0,104	0,074	0,025	0,018	0,007	0,145

Tabuľka 2: Smerodajné odchýlky u mužov v jednotlivých vekových skupinách pri jednotlivých meraniach na pravej a na ľavej ACC

	MUŽI					M Total	Grand Total
Values	30-40	41-50	51-60	61-70	71+		
StdDev of l.dx.I.	0,128	0,064	0,033	0,041	0,010	0,157	0,153
StdDev of l.dx.II.	0,126	0,062	0,033	0,035	0,009	0,155	0,150
StdDev of l.dx.III.	0,127	0,061	0,033	0,036	0,008	0,152	0,148
StdDev of l.dx.P.	0,127	0,062	0,031	0,046	0,008	0,156	0,151
StdDev of l.sin.I.	0,118	0,062	0,027	0,042	0,007	0,151	0,148
StdDev of l.sin.II.	0,119	0,065	0,029	0,040	0,008	0,150	0,146
StdDev of l.sin.III.	0,119	0,064	0,031	0,041	0,011	0,149	0,145
StdDev of l.sin.P	0,119	0,063	0,028	0,040	0,008	0,150	0,147

Z vypočítaných hodnôt môžeme konštatovať že najväčšia variabilita je medzi 30-40 ročných mužov aj žien. Variabilita vekom postupne klesá.

Na základe vypočítania percentilov sa stanovili hodnoty rozsahu normálnych hodnôt podľa tabuľky č. 3.

Tabuľka 3: Vybrané percentily súboru zdravých respondentov a podsúborov podľa veku a pohlavia podľa vekových skupín a podľa pohlavia a podľa pravej a ľavej strany

Percentily	všetky vekové skupiny spolu		Percentily	všetky pacienti vo vekovej skupine 30-40 r.		Percentily	všetky pacienti vo vekovej skupine 41-50 r.		Percentily	všetky pacienti vo vekovej skupine 51-60 r.	
	l.dx.	l.sin.		l.dx.	l.sin.		l.dx.	l.sin.		l.dx.	l.sin.
0,05	0,297	0,328	0,05	0,23335	0,24675	0,05	0,6107	0,56525	0,05	0,63845	0,64915
0,1	0,442	0,428	0,1	0,2548	0,264	0,1	0,6347	0,6239	0,1	0,6453	0,6539
0,5	0,713	0,707	0,5	0,4425	0,433	0,5	0,7	0,694	0,5	0,6925	0,6935
0,9	0,808	0,801	0,9	0,5534	0,5514	0,9	0,745	0,7387	0,9	0,7207	0,7222
0,95	0,815	0,805	0,95	0,57535	0,5605	0,95	0,756	0,75765	0,95	0,72865	0,72675
Percentily	všetky vekové kategórie žien spolu		Percentily	všetky ženy vo vekovej skupine 30-40 rokov		Percentily	všetky ženy vo vekovej skupine 41-50 rokov		Percentily	všetky ženy vo vekovej skupine 51-60 rokov	
	l.dx.	l.sin.		l.dx.	l.sin.		l.dx.	l.sin.		l.dx.	l.sin.
0,05	0,382	0,375	0,05	0,2504	0,2504	0,05	0,63155	0,56865	0,05	0,6539	0,65365
0,1	0,433	0,404	0,1	0,2578	0,2578	0,1	0,6422	0,6271	0,1	0,6672	0,6579
0,5	0,715	0,707	0,5	0,442	0,442	0,5	0,6995	0,694	0,5	0,706	0,6985
0,9	0,802	0,800	0,9	0,5622	0,5622	0,9	0,7429	0,7429	0,9	0,7249	0,7228
0,95	0,809	0,805	0,95	0,5854	0,5854	0,95	0,752	0,75305	0,95	0,72805	0,72575
Percentily	všetky vekové kategórie mužov spolu		Percentily	všetky muži vo vekovej kategórii 30-40 r.		Percentily	všetky muži vo vekovej kategórii 41-50 r.		Percentily	všetky muži vo vekovej kategórii 51-60 r.	
	l.dx.	l.sin.		l.dx.	l.sin.		l.dx.	l.sin.		l.dx.	l.sin.
0,05	0,2812	0,2846	0,05	0,217	0,239	0,05	0,592	0,59675	0,05	0,6388	0,6456
0,1	0,447	0,4584	0,1	0,250	0,255	0,1	0,6225	0,6235	0,1	0,6432	0,6508

0,5	0,709	0,703	0,5	0,443	0,446	0,5	0,7005	0,696	0,5	0,682	0,689
0,9	0,8142	0,8018	0,9	0,550	0,537	0,9	0,74	0,7335	0,9	0,7118	0,7106
0,95	0,816	0,8093	0,95	0,556	0,552	0,95	0,7625	0,74475	0,95	0,7222	0,7288

Percentily	všetky pacienti vo vekovej skupine 61-70 r.		Percentily	všetky pacienti vo vekovej skupine viac ako 70 r.	
	l.dx.	l.sin		l.dx.	l.sin
0,05	0,735	0,7294	0,05	0,79115	0,7903
0,1	0,7359	0,7396	0,1	0,7944	0,7923
0,5	0,766	0,759	0,5	0,808	0,803
0,9	0,7935	0,7845	0,9	0,816	0,8114
0,95	0,8334	0,7901	0,95	0,816	0,81285
Percentily	všetky ženy vo vekovej skupine 61-70 rokov		Percentily	všetky ženy vo vekovej skupine viac ako 70 rokov	
	l.dx.	l.sin		l.dx.	l.sin
0,05	0,735	0,7352	0,05	0,7905	0,791
0,1	0,739	0,7404	0,1	0,791	0,792
0,5	0,765	0,764	0,5	0,804	0,803
0,9	0,7838	0,7824	0,9	0,811	0,808
0,95	0,7958	0,7861	0,95	0,812	0,8125
Percentily	všetky muži vo vekovej kategórii 61-70 r.		Percentily	všetky muži vo vekovej kategórii viac ako 70 r.	
	l.dx.	l.sin		l.dx.	l.sin
0,05	0,7186	0,7162	0,05	0,7878	0,78425
0,1	0,7444	0,7404	0,1	0,795	0,7881
0,5	0,777	0,758	0,5	0,8105	0,8015
0,9	0,8298	0,787	0,9	0,816	0,8114
0,95	0,8619	0,8121	0,95	0,81635	0,81235

Za normálne hodnoty môžeme považovať hodnoty medzi 10. a 90. percentilom (v prísnejšom pohľade) a vo voľnejšom pohľade medzi 5. a 95. percentilom. Hodnoty 5. resp. 10. percentilu znamenajú, že nameraná hodnota IMT (vpravo a vľavo) by nemala byť menšia ako 5. resp. 10. percentil, a na druhej strane by nemala byť väčšia ako 90. resp. 95. percentil. (Chajdiak, 2003, Chajdiak, 2004).

5. Diskusia

V súbore 144 zdravých respondentov vo vekovej skupine ≥ 30 r. sme stanovili norm. hodnoty hrúbky IMT. Celkove sme vykonali 864 vyšetrení CIMT, z ktorých sme vypočítali smerodajné odchýlky IMT. Najväčšiu variabilitu CIMT sme zistili u 30-40 ročných, vekom variabilita klesá. Zistili sme, že narastanie hrúbky intima-médie vykazuje významnú závislosť od veku pacienta. Výsledky nameraných maximálnych priemerných hodnôt nepresahujú v našom súbore 0.85 mm a sú v súlade s literárnymi údajmi (12,13,14,15).

Normálne hodnoty sme stanovili na základe vypočítania 5., 10., 90. a 95-tého percentilu. Za normálne hodnoty môžeme považovať hodnoty medzi 10. a 90. percentilom (v prísnejšom pohľade) a vo voľnejšom pohľade medzi 5. a 95. percentilom. (Chajdiak 2003, Chajdiak, 2004). Hodnoty 5./10. percentilu znamenajú, že nameraná hodnota CIMT (dx. a

sin.) by nemala byť menšia ako 5./10. percentil, a na druhej strane by nemala byť väčšia ako 90./95. percentil (tab. č. 3.)

Záver

Záverom môžeme konštatovať, že normálne hodnoty vypočítané pomocou MS Office, excel sú v súlade s lierálnymi údajmi v hodnotení v percentiloch, a nepresahujú maximálne hodnoty 0,85 mm.

Odporúčame zavedenie metódy merania CIMT do každodennej klinickej praxe pri hodnotení prítomnosti rizika CHOS.

Pri vyšetrení CIMT odporúčame postupovať podľa MIMTC z r. 2004 a 2006 a na vyšetrenie používať automatický alebo semiautomatický systém merania CIMT.

Metóda je jednoduchá, kedykoľvek opakovateľná, nebolestivá, bez rizika - odporúčame využívať pri sledovaní progresie AS postihnutia cievneho systému.

Literatúra

- [1] CSIBA, L.: Az ultrahanggal mért arteria carotis intima-media vastagság: a különböző gyógyszerek hatása az atherosclerosis progressiójára, Orv. hetil. 2005. 146 évf. 23. sz. 1245–1249. ISSN 0030-6002
- [2] BOTS, M.L. et al.: *Carotid Intima-Media Thickness Measurements in Intervention Studies. Design Options, Progression Rates, and Sample Size Considerations: A Point of View.* Stroke. 2003;34:2985-2994.
- [3] CSÁNYI A.: Magyar Stroke Társaság és a Magyar Neuroszonológiai Társaság VII. Konferenciája, Az akut stroke ellátás Nyugat-és Közép/Kelet - Európában Eger, szept. 29.-okt.1, Hungary, 2005.
- [4] DE GROOT, R. et al.: Measurement of carotid intima-media thickness to assess progression an regression of atherosclerosis. Nature Clinical Practice Cardiovascular Medicine 5, 2008, 280-288. ISSN:1759-5002
- [5] CHAJDIK, J. Štatistické úlohy a ich riešenie v exceli. Bratislava:STATIS, 2005, ISBN 80-85659-39-5, s. 11-75.
- [6] CHAJDIK, J. Štatistika Jednoducho. Bratislava:STATIS, 2003, ISBN 80-85659-28-5, s. 9-91.
- [7] IMT LAB WWSM. Corporate MKT. Fy BIOESAOTE Italiano. Propagačné materiálny. Vienna:2005
- [8] LINHART, A., SKALICKÁ, L., GOLÁNY, L., CHARALAMPIDI, K., ASCHERMANN M.:Stratifikace rizika a hodnocení efektu léčby stanovením tloušťky intimy a medie karotických tepen, Postgraduální medicína, 2003, 5, č. 3, str. 296-301. ISSN: 1210-7883.
- [9] MADARÁSZ, Š.:Carotid Intima-Media Thickness - a reliable Marker of Atherosclerosis. Artery Research (2008), XX, 1-8, DOI:10.1016/J.ARTERS.2008.02.003. ISSN: 1872-9312.
- [10] MADARÁSZ, Š.: Demografické údaje o chorobách obehovej sústavy a cievnych mozgových chorobách v Slovenskej republike. Medicina Militaris Slovaca, odborný časopis zdravotníckej služby ozbrojených síl Slovenskej republiky, ročník 10, 1/2008, str. 43-47. Vydavateľské oddelenie komunikačného odboru MOSR, ISSN 1335-5139.

- [11] PIGNOLI P, TREMOLI E, POLI A, ORESTE P, PAOLETTI R.: Intimal plus medial thickness of the arterial wall: a direct measurement with ultrasound imaging, *Circulation* 74, No. 6, 1399-1406, 1986. ISSN: 0009-7322.
- [12] SPÁČIL J, PETRÁŠEK J, ŠPERL M.: sonografie karotid a měření intimomediální tloušťky jako ukazatelé rizika generalizované ateroskleriózy, *Časopis lékaře českých*, 141, 2002, č. 4, str. 118-121. ISSN 0008-7335.
- [13] ŠIMON, J.: Perspektivy preventivní kardiologie aneb co dále. *Cor Vasa* 2008;50(4):145–146. ISSN 0010-8650.
- [14] ŠKOLOUDÍK D, ŠKODA O, BAR M, BROZMAN M, VÁCLAVÍK D.: *Neurosonologie*, Galén, Prvé vydanie, 2003, Praha 5, str. 52-53, ISBN 80-7262-245-5.
- [15] ŠKOLOUDÍK, D. : *Neuro3*, 2004 10:R01
- [16] TOUBOUL, P.J, HENNERICI, M.G., MEAIRS, S. ET AL.: Mannheim Intima-Media Thickness Consensus on Behalf of the Advisory Board of the 3rd Watching the Risk Symposium 2004, 13th European Stroke Conference. Mannheim, Germany, May 14, 2004. *Cerebrovasc Dis* 2004;18:346-349, ISSN 1015-9770.
- [17] 2006 MANNHEIM INTIMA-MEDIA THICKNESS CONSENSUS - An Update on Behalf of the Advisory Board of the 3rd and 4th Watching the Risk Symposium 13th and 15th European Stroke Conferences, Mannheim, Germany, 2004, and Brussels, Belgium, 2006, *Cerebrovasc Dis* 2007;23:75–80. ISSN 1015-9770.

Adresa autorov:

Štefan Madarász, MUDr., PhD.
Neurologická klinika, ÚVN SNP FN
Ružomberok a
Inštitút fyzioterapie, balneológie a liečebnej
rehabilitácie, Piešťany, UCM Trnava
madaraszs@gmail.com

Jozef Chajdiak, Doc., Ing., CSc.
ÚM STU
Bratislava
chajdiak@statis.biz

Hodnotenie vzťahu patologickej hrúbky intima-medie k ostatným rizikovým faktorom

Assessment of relation between pathological thickness of intima-medie and other danger factors

Madarász Štefan, Chajdiak Jozef

Abstract: We know, that in case of presence of one danger factor with activ searching a possibility, that we can find others, is high, what has a huge meaning for patient in primary prevention.

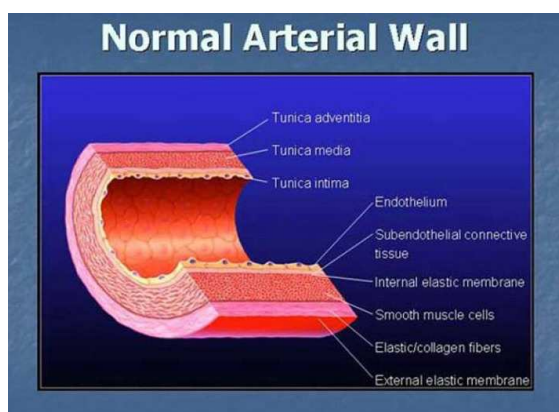
The authors were finding and statistically found out by means of programmatic bundle MS Office, Excel a mutual correlation between selected danger factors and pathological thickness of intima-medie.

Key words: danger factors, MS-office, excel, correlative matrix

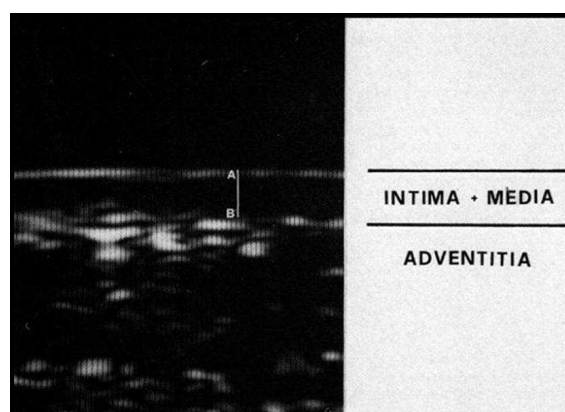
Kľúčové slová: rizikové faktory, MS-office, excel , korelačná matica.

1. Úvod

Najčastejšou príčinou chorôb obehovej sústavy, ktoré vedú k závažným komplikáciám u pacienta je ateroskleróza. Progresia procesu sa mnoho krát dá zistiť a sledovať meraním intima-mediálneho komplexu pomocou ultrazvuku [16]. Intima-media (IMT) sa dá definovať ako vnútorná časť cievnej steny čo sa prejavuje v ultrazvukovom obraze ako šírka echogenného rozhrania medzi lúmenom a intimou a echogenného rozhrania medzi médiou a adventíciou cievnej steny [1] [6] [7] [8] [13] [14] [15].



Obrázok 1: Histologická skladba normálnej arteriálnej steny podľa Roosa [14].



Obrázok 2: B- obraz cievnej steny zobrazený pignolim a spol. Vzďialenosť medzi dvomi dvomi hyperechogennými líniami A+B zobrazuje intima-médiu [13].

Vieme, že v prípade prítomnosti jedného rizikového faktoru u pacienta pri aktívnom hľadaní pravdepodobnosť, že sa nájdu aj ďalšie, je vysoká, čo pre pacienta v rámci primárnej prevencie má veľký význam. [9] [12].

2. Cieľ práce

Autori sa zamerali na zistenie vzájomnej korelácie medzi vybranými rizikovými faktormi CHOS a patologickou hrúbkou intima-médie.

Hľadali a štatisticky vyhodnotili pomocou programového balíka MS Office, excel, vzájomnú koreláciu medzi vybranými rizikovými faktormi a patologickou hrúbkou intima-médie.

2. Materiál a metóda

Do súboru bolo zaradených 85 pacientov s patologickým IMT. Súbor tvorilo 51 mužov a 34 žien. Vekovým priemerom v súbore bol 50,6 rokov, vekový priemer mužov bol 48,6 rokov a vekový priemer žien 53,3 rokov. Minimálny vek v súbore bol 28 rokov, a maximálny 86 rokov. V skupine mužov minimálny vek bol 32 rokov, a maximálny 83 rokov, v skupine žien minimálny vek bol 28 rokov a maximálny 80 rokov.

Sledovali a štatisticky sme porovnávali výskyt pacientov s hypercholesterolémiou, kde z 85 pacientov 36, t.j. 42,4% pacientov malo zvýšenie celkového cholesterolu. Z 36 pacientov 20 bolo mužov, t.j. (23,6% z celkového počtu pacientov) a 16 žien t.j. (18,8% z celkového počtu). Ako ďalší rizikový faktor sme si vybrali fajčenie. Zo súboru 85 pacientov s patologickým IMT, 37 pacientov fajčilo, čo je 43,5%, z toho bolo 19 mužov (23,4% z celého súboru), a 18 žien, čo je 21,2% zo súboru 85 pacientov s patologickým IMT. Ďalším rizikovým faktorom bola nadváha, (hodnotili sme ju podľa BMI indexu), ktorú sme našli u 59 pacientov, čo je (69,4%). S nadváhou trpelo zo súboru 40 mužov (47,1%) a 19 žien (22,3%). Vysokým systolickým + diastolickým tlakom trpelo 18 pacientov, čo je 21,2 pacientov, z toho bolo 14 mužov a 4 ženy t.j. (16,5% / 4,7%). Zvýšený systolický tlak malo 28 pacientov (32,9%), 16 mužov (18,8%) a 12 žien (14,1%), a len diastolický tlak mali zvýšený 24 pacienti (28,2%), z toho 19 muži (22,4%) a 5 žien (5,8%). Ako jeden z najvýznamnejších rizikových faktorov okrem hypertenzie je diabetes mellitus, čo sme našli v súbore u 27 pacientov (31,76%), z toho 16 mužov (18,8%) a 11 žien (12,9%). Jedným z novo uznaných rizikových faktorov je zvýšenie arteriálneho stiffnessu, čo sa vyjadruje meraním augmentačného indexu (Aix) a rýchlosťou šírenia pulzovej vlny aorty (PWV ao). V našom súbore zvýšenie Aix sa našlo u 31 pacientov (36,5%), a to u 12 mužov (14,1%) a 19 žien (22,4%), a patologické PWV ao u 33 pacientov (38,8%), u 11 mužov (12,9%) a u 22 žien (25,9%).

Vyšetrenie IMT sme vykonali na celotelovom ultrasonografickom prístroji Aloka 4000 s 12 MHz multifokálnou a multiflekvenčnou lineárnou sondou s vysokým rozlíšením.

Pri vyšetrení a zobrazení intima-medie sa dodržali odporúčania Mannheim Intima-Media Tthickness consensus z r. 2004 [16].

Na monitore zobrazený obraz intima-médie sa archivoval v jpeg formáte na archivačnom médiu, a následne sa merala hrúbka intima média poloaťomatizovanou metódou pomocou softvéru IMT LAB ESAOTE verzia 1.1 na archivovaných snímkach s presnosťou na tisícinu milimetra. [11].

Miera nadváhy sa hodnotila pomocou BMI indexu. Biochemické vyšetrenia sa robili podľa štandardných metód v biochemickom laboratóriu. Krvný tlak sa meral u každého pacienta podľa štandardných podmienok digitálnym tlakomerom Tensoval duo control.

Arteriálnej stiffness sa vyšetroval prístrojom Arteriograf Tensiomed podľa odporúčania výrobcu s používaním normatívnych hodnôt udávaných výrobcom[17]

Štatistické spracovanie údajov sa robilo pomocou programového balíka MS Office, Excel, funkciou CORREL sa hodnotil korelačný koeficient medzi dvoma množinami, a zisťoval sa vzťah prítomných rizikových faktorov u pacienta k nálezu patologického IMT. [4] [5]

4. Výsledky

Korelácie medzi IMT vpravo a vľavo a ostatnými rizikovými faktormi boli vypočítané pomocou funkcie CORREL programového vybavenia MS Office, Excel.

KORELÁCIE MEDZI IMT VPRAVO A VĽAVO A OSTATNÝMI RIZIKOVÝMI FAKTORMI			
I.dx.	VEK	I.sin.	VEK
	0,761		0,756
I.dx.	PWV	I.sin.	PWV
	0,446		0,439
I.dx.	SyTK	I.sin.	SyTK
	0,324		0,311
I.dx.	DiTK	I.sin.	DiTK
	0,232		0,225
I.dx.	AIX%	I.sin.	AIX%
	0,221		0,219
I.dx.	BMI	I.sin.	BMI
	0,214		0,207
I.dx.	DM	I.sin.	DM
	0,039		0,038

Tabuľka 1. vyjadruje vzťah prítomných rizikových faktorov k patologickej IMT v súbore 85 pacientov

Vytvorená korelačná matica dáva možnosť na porovnanie vzájomných vzťahov medzi výskytom jednotlivých sledovaných rizikových faktorov v našom súbore.

	VEK	IMT I.dx.	IMT I.sin.	CHOL	BMI	SyTK	DiTK	AIX	PWS
VEK	1								
IMT I.dx.	0,761	1							
IMT I.sin.	0,756	0,998	1						
CHOL	0,045	0,206	0,216	1					
BMI	0,094	0,214	0,207	0,150	1				

SyTK	0,220	0,324	0,311	0,076	0,494	1			
DiTK	0,219	0,232	0,225	0,067	0,472	0,795	1		
AIX	0,433	0,221	0,219	-0,061	0,054	0,135	0,165	1	
PWS	0,506	0,446	0,439	0,163	0,362	0,272	0,218	0,598	1

Tabuľka 2. Výstup korelačnej matice závislosti medzi pravým a ľavým IMT s cholesterolom, BMI, SyTK, DiTK, Aix a PWV

5. Diskusia

Z hodnôt korelačnej matice vidíme, že najvyššia miera vzájomnej závislosti je medzi pravým a ľavým IMT (0,998). Nasleduje trochu nižšia závislosť medzi SyTK a DiTK (0,795), medzi vekom a IMT (0,761 a 0,756). Stredná miera závislosti je medzi BMI a SyTK (0,494), medzi BMI a DiTK (0,472), medzi Aix a PWV (0,598). Slabá závislosť je medzi cholesterolom a PWV (0,163), a cholesterolom a Aix (-0,061, napriek veľmi nízkej hodnoty je záporná hodnota prekvapivá). Ostatné hodnoty reprezentujú slabšiu mieru vzájomnej závislosti. [4] [5].

6. Záver

V úvode predpokladaná hypotéza sa potvrdila t.j. že v prítomnosti patologickej hrúbky intima média pravdepodobne u pacientov sa zistia aj ďalšie rizikové faktory. U väčšiny sledovaných pacientov sme zistili aj ďalšie prítomné rizikové faktory. Má to nesmierny veľký význam v primárnej prevencii ktorá zahŕňa v sebe aktívne vyhľadávanie rizikových pacientov a ich liečbu. [2] [3] [10] [11].

Je to aktívne vyhľadávanie jednotlivcov bez klinických ťažkostí s prítomnými rizikovými faktormi pomocou screeningových preventívnych vyšetrení ako je napríklad aj meranie intima-médie, s ich následnou adekvátnou cielehou liečbou. Taktiež má význam aj u pacientov s klinickými ťažkosťami ale bez adekvátnej liečby[10] [11].

Literatúra

- [1]BOTS, M.L. et al.: *Carotid Intima-Media Thickness Measurements in Intervention Studies*. Design Options, Progression Rates, and Sample Size Considerations: A Point of View. Stroke. 2003;34:2985-2994.
- [2]CÍFKOVÁ, R. Epidemiologie KV nemocí v ČR - výsledek prevence, včasné diagnostiky a efektivní léčby, pohled na roli primární péče, Institut klinické a experimentální medicíny Praha.in.: <http://www.vpl.sk/cps/rde/xbcr/sprlsk/2.Cifkova.pdf>.
- [3]CÍFKOVÁ, R. et al. 2005. Prevence kardiovaskulárních onemocnění v dospělém věku. Společné doporučení českých odborných společností, Supplementum Cor Vasa: 2005;47(9):3-14. ISSN 0010-8650
- [4]CHAJDIK, J. Štatistické úlohy a ich riešenie v exceli. Bratislava:STATIS, 2005, ISBN 80-85659-39-5, s. 11-75.
- [5]CHAJDIK, J. Štatistika Jednoducho. Bratislava:STATIS, 2003, ISBN 80-85659-28-5, s. 9-91.

- [6]CSÁNYI A.: Magyar Stroke Társaság és a Magyar Neuroszonológiai Társaság VII. Konferenciája, Az akut stroke ellátás Nyugat-és Közép/Kelet - Európában Eger, szept. 29.-okt.1, Hungary, 2005.
- [7]DE GROOT, R. et al.: Measurement of carotid intima-media thickness to assess progression an regression of atherosclerosis. *Nature Clinical Practice Cardiovascular Medicine* 5, 2008, 280-288. ISSN:1759-5002
- [8]KAPPERT, A.: 1987 Angiológia. Učebnica a atlas. Choroby tepien, žíl, vlásoknic lymfatických ciev. Martin : Osveta, 1987. SÚKK 1721/i.-86.
- [9]MADARÁSZ, Š.:Carotid Intima-Media Thickness - a reliable Marker of Atherosclerosis. *Artery Research* (2008), XX, 1-8, DOI:10.1016/J.ARTERS.2008.02.003. ISSN: 1872-9312.
- [10]MADARÁSZ, Š.: Primárna prevencia cerebrovaskulárnych ochorení. *Medicina Militaris Slovaca*, odborný časopis zdravotníckej služby ozbrojených síl Slovenskej republiky, ročník 10, 1/2008, str. 43-47. Vydavateľské oddelenie komunikačného odboru MOSR, ISSN 1335-5139
- [11]MADARÁSZ, Š. 2009. Zavedenie merania intima-médie ako ukazovateľa efektivity liečby a prevencie. Dizertačná práca. Bratislava:2009. s. 152.
- [12]NAGY, Z. et al. 2002. A cerebrovaskuláris betegségek megelőzése, diagnosztikája, akut ellátása és korai rehabilitációja, Tényekre támaszkodó ajánlások. *Agyérbetegségek* 2002;8(Suppl 1):2-18. ISSN:1217-7962.
- [13]PIGNOLI P, TREMOLI E, POLI A, ORESTE P, PAOLETTI R.: Intimal plus medial thickness of the arterial wall: a direct measurement with ultrasound imaging, *Circulation* 74, No. 6, 1399-1406, 1986. ISSN: 0009-7322.
- [14]ROSS, R. 1990. The pathogenesis of atherosclerosis: a perspective for the 1990s, *Nature* 362, p. 801-809. ISSN : 0028-0836.
- [15]SCHULTE-ALTEDORNEBURG, G. et al. 2001. Accuracy of In Vivo Carotid B-Mode Ultrasound Compared With Pathological Analysis, Intima-Media Thickening, Lumen Diameter, and Cross-Sectional Area, *Stroke* : 2001;32:1520. ISSN:00392499
- [16]TOUBOUL, P.J, HENNERICI, M.G., MEAIRS, S. ET AL.: Mannheim Intima-Media Thickness Consensus on Behalf of the Advisory Board of the 3rd Watching the Risk Symposium 2004,13th European Stroke Conference. Mannheim, Germany, May 14, 2004. *Cerebrovasc Dis* 2004;18:346-349, ISSN 1015-9770.
- [17]TENSIO MED ARTERIOGRÁF -5-00, V 1.9.1. Felhasználói kézikönyv. Budapest : 2005.

Adresa autorov:

Štefan Madarász, MUDr., PhD.
Neurologická klinika, ÚVN SNP FN
Ružomberok a
Inštitút fyzioterapie, balneológie a liečebnej
rehabilitácie, Piešťany, UCM Trnava
madaraszs@gmail.com

Jozef Chajdiak, Doc., Ing., CSc.
ÚM STU
Bratislava
chajdiak@statis.biz

OBSAH

	Úvod	1
Luha J.	Základné oblasti a úlohy medicínskej štatistiky	2
Luha J.	Metodologické zásady záznamu dát z rozličných oblastí medicíny a zásady ich kontroly	16
Chajdiak J.	Kontrola súborov vstupných diskretných hodnôt pomocou podsystemu Pivot Table systému Excel verzie 2007	39
Chajdiak J.	Súbory údajov pre MEDSTAT2010	47
Letkovičová M., Zach H.	Neurónové siete v ekológii	51
Letkovičová M., Zach H.	Priestorová štúdia rozmiestnenia rýchlej zdravotnej pomoci	63
Luha J.	2x2 Kontingenčné tabuľky a diagnostické testy	71
Böhmer D., Braxatorisová T., Luha J., Malová J., Repiská V., Vojtaššák J.	Morfológia a karyotypy databázy Abortus	79
Páleník V., Kvetan V., Mlýnek M., Radvanský M.	Dlhodobá prognóza príjmov a výdavkov slovenského zdravotného systému s analýzou senzitivnosti	89
Páleník V., Kvetan V.	Analýza determinantov samo hodnotenia zdravotného stavu a využitia zdravotnej starostlivosti	107
Radvanský M.	Dlhodobá starostlivosť na Slovensku	117
Moravčík T.	Inteligentné technológie – genetický algoritmus	126
Stankovičová I., Suchánek M., Chajdiak J.	Využitie logistickej regresie pri modelovaní chorobných stavov v gynekológii	129
Madarász Š., Chajdiak J.	Výpočet normatívnych hodnôt intima-medie na artérii carotis communis	136
Madarász Š., Chajdiak J.	Hodnotenie vzťahu patologickej hrúbky intima-medie k ostatným rizikovým faktorom	144
	Obsah	149
	Contents	150

CONTENTS

	Introduction	1
Luha J.	Basic sections and tasks of medical statistics	2
Luha J.	Metodological principles of entering data from different sections of medicine and principles of their checking	16
Chajdiak J.	Verification file input discrete values by subsystem Pivot Table of the system Excel (version 2007)	39
Chajdiak J.	Input file for MEDSTAT 2010	47
Letkovičová M., Zach H.	Neural networks in ecology	51
Letkovičová M., Zach H.	Space study of allocation of health rescue services	63
Luha J.	2x2 Contingency tables and diagnostic tests	71
Böhmer D., Braxatorisová T., Luha J., Malová J., Repiská V., Vojtaššák J.	Morphology and Karyotypes of database Abortus	79
Páleník V., Kvetan V., Mlýnek M., Radvanský M.	Long term forecast of the revenue and expense of the Slovak health care system and analysis his sensibility	89
Páleník V., Kvetan V.	Analysis of the determination of self-assessed health and healthcare utilisation	107
Radvanský M.	Long-Term Care in Slovakia	117
Moravčík T..	Inteligent Technologies – Genetic algorithm	126
Stankovičová I., Suchánek M., Chajdiak J.	Application of logistic regression for modelling of pathological phenomenon in gynaecology	129
Madarász Š., Chajdiak J.	Calculation prescriptively values intima-medie on the carotis communis artery	136
Madarász Š., Chajdiak J.	Assessment of relation between pathological thickness of intima-medie and other danger factors	144
	Obsah	149
	Contents	150

Pokyny pre autorov

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

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

Rozsah: Maximálny rozsah príspevku je 6 strán.

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

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

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

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

Vynechať riadok

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

Vynechať riadok

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

Vynechať riadok

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

Vynechať riadok

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

Vynechať riadok

Vlastný text príspevku v členení:

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

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

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

[1] Písať podľa normy STN ISO 690

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

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

Meno1 Priezvisko1, tituly1
Ulica1
970 00 Mesto1
meno1.priezvisko1@mail.sk

Meno2 Priezvisko2, tituly2
Ulica2
970 00 Mesto2
meno2.priezvisko2@mail.sk

FORUM STATISTICUM SLOVACUM

vedecký recenzovaný časopis Slovenskej štatistickej a demografickej spoločnosti

Vydavateľ

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

Redakcia

Miletičova 3
824 67 Bratislava 24
Slovenská republika

Fax

02/39004009

e-mail

chajdiak@statis.biz
jan.luha@fmed.uniba.sk

Registráciu vykonalo

Ministerstvo kultúry Slovenskej republiky

Registračné číslo

3416/2005

Evidenčné číslo

EV 3287/09

Tematická skupina

B1

Dátum registrácie

22. 7. 2005

Objednávky

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

ISSN 1336-7420

Redakčná rada

RNDr. Peter Mach – *predseda*

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

RNDr. Ján Luha, CSc. – *tajomník*

členovia:

Ing. František Bernadič
Mgr. Branislav Bleha, PhD.
Ing. Mikuláš Cár, CSc.
Ing. Ján Cuper
Ing. Pavel Flák, DrSc.
Ing. Edita Holíčková
Doc. RNDr. Ivan Janiga, CSc.
Ing. Anna Janusová
Doc. RNDr. PaedDr. Stanislav Katina, PhD.
Prof. RNDr. Jozef Komorník, DrSc.
RNDr. Samuel Koróny, PhD.
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.
Prof. RNDr. Beata Stehlíková, CSc.
Prof. RNDr. Anna Tirpáková, 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

VI.

Číslo

1/2010

Cena výtlačku 20 EUR

Ročné predplatné 80 EUR