

Štatistický úrad Slovenskej republiky
The Statistical Office of the Slovak Republic

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

vedecký časopis/scientific journal

2/2022
ročník 32



ŠTATISTICKÝ
ÚRAD
SLOVENSKEJ
REPUBLIKY

ISSN 1339-6854 (online)
ISSN 1210-1095 (tlačené vydanie)

SLOVENSKÁ ŠTATISTIKA A DEMOGRAFIA

Recenzovaný vedecký časopis založený v roku 1991. Od roku 2014 jednotlivé čísla časopisu zverejňujeme aj v elektronickej podobe na ssad.statistics.sk. Názory autorov článkov sa nemusia zhodovať s názormi vydavateľa.

Zahraniční poradcovia/Foreign Consultants

Gabriela Czanner

University of Liverpool
Veľká Británia/United Kingdom

Jitka Langhamrová

Vysoká škola ekonomická v Praze
University of Economics in Prague
Česká republika/Czech Republic

Estefanía Mourelle Espasandín

Universidade da Coruña
Španielsko/Spain

Michaela Potančoková

Joint Research Centre,
European Commission
Taliansko/Italy

Hana Řezanková

Vysoká škola ekonomická v Praze
University of Economics in Prague
Česká republika/Czech Republic

Milan Stehlík

Institute of Statistics, University of Valparaíso
Čile/Chile
Johannes Kepler University Linz
Rakúsko/Austria

Výkonná redaktorka/Executive Editor

Silvia Hudecová

Jazykové redaktorky/Language Editors

Slovenský jazyk/Slovak Language

Silvia Duchková

Anglický jazyk/English Language

Andrea Okenková

SLOVAK STATISTICS AND DEMOGRAPHY

The scientific peer-reviewed journal founded in 1991. From 2014 individual copies of the journal are available to readers in electronic form at the website ssad.statistics.sk. The opinions of the authors do not necessarily correlate with the opinions of the publisher.

Redakčná rada/Editorial Board

Ľudmila Ivančíková

(predsedníčka/chairwoman)
Štatistický úrad SR
Statistical Office of the SR

Mikuláš Cár

Slovenská štatistická a demografická spoločnosť
Slovak Statistical and Demographic Society

Helena Glaser-Opitzová

Štatistický úrad SR
Statistical Office of the SR

Ján Haluška

INFOSTAT Bratislava

Iveta Stankovičová

Univerzita Komenského v Bratislave
Comenius University in Bratislava

Erik Šoltés

Ekonomická univerzita v Bratislave
University of Economics in Bratislava

Pavol Tišliar

Univerzita Cyrila a Metoda v Trnave
University of Ss. Cyril and Methodius in Trnava
Masarykova univerzita
Masaryk University

Boris Vaňo

INFOSTAT - Výskumné demografické centrum
INFOSTAT - Demographic Research Centre

Adresa redakcie/Address of Editorial Office

Slovenská štatistika a demografia
Štatistický úrad SR
Lamačská cesta 3/C, 840 05 Bratislava 45
Slovenská republika

E-mailová adresa/E-mail address

SSaD@statistics.sk

ssad.statistics.sk
www.statistics.sk

OBSAH/CONTENTS

I. VEDECKÉ ČLÁNKY/SCIENTIFIC ARTICLES

Branislav ŠPROCHA 3
SOBÁŠNOSŤ SLOBODNÝCH NA SLOVENSKU V PANDEMICKOM ROKU 2020
NUPTIALITY OF SINGLE PERSONS IN SLOVAKIA IN THE PANDEMIC YEAR 2020

Mikuláš CÁR 19
HLADANIE ÚČINNÝCH METÓD A NÁSTROJOV NA HODNOTENIE EPIDEMICKEJ
SITUÁCIE
SEARCHING FOR EFFECTIVE METHODS AND TOOLS TO ASSESS
THE EPIDEMIOLOGICAL SITUATION

Milan TEREK 34
ANALÝZA KAUZÁLNYCH VZŤAHOV MEDZI PREMENNÝMI
ANALYSIS OF CAUSAL RELATIONS BETWEEN VARIABLES

Boris VAŇO 52
HODNOTENIE KVALITY ADMINISTRATÍVNYCH ÚDAJOV PRE POTREBY
ŠTATISTICKÉHO ÚRADU SR
QUALITY EVALUATION OF ADMINISTRATIVE DATA FOR THE NEEDS OF THE
STATISTICAL OFFICE OF THE SLOVAK REPUBLIC

II. INFORMATÍVNE ČLÁNKY, NÁZORY, RECENZIE, ROZHOVORY, INFORMÁCIE/ INFORMATIVE ARTICLES, OPINIONS, REVIEWS, INTERVIEWS, INFORMATION

Andrea KADEROVÁ 67
Michal Páleš, František Slaninka
TEÓRIA RIZIKA V POISTENÍ – RIEŠENÉ PRÍKLADY V JAZYKU R A MAXIMA
Michal Páleš, František Slaninka
RISK THEORY IN INSURANCE – SOLVED EXAMPLES IN LANGUAGE R AND
MAXIMA
Recenzia publikácie/Review of Publication

III. PRIPRAVUJEME/COMING SOON 69

Branislav ŠPROCHA
INFOSTAT – Výskumné demografické centrum
Centrum spoločenských a psychologických vied SAV

SOBÁŠNOSŤ SLOBODNÝCH NA SLOVENSKU V PANDEMICKOM ROKU 2020¹

NUPTIALITY OF SINGLE PERSONS IN SLOVAKIA IN THE PANDEMIC YEAR 2020

ABSTRAKT

Celosvetová pandémia spojená s ochorením COVID-19 výrazným spôsobom zasiahla nielen zdravotný stav a úmrtnostné pomery populácie Slovenska, ale v mnohých smeroch paralyzovala bežný chod našej spoločnosti. Prijímané hygienické a epidemiologické opatrenia so snahou zamedziť prenos a šírenie ochorenia, ako aj ekonomická neistota a psychologický efekt obavy z nepoznanej situácie mohli negatívne vplývať aj na rozhodovacie procesy mladých ľudí pri realizácii niektorých dlhodobých rozhodnutí. Do tejto skupiny je možné zaradiť aj rozhodovanie o uzavretí manželstva. Cieľom príspevku je prostredníctvom dostupných údajov z roku 2020 analyzovať najmä intenzitu a časovanie procesu sobášnosti slobodných na Slovensku a jej prípadné zmeny, ktoré podmienila práve špecifická situácia v pandemickom roku 2020.

ABSTRACT

The global pandemic associated with the COVID-19 disease significantly affected not only the health status and mortality rates of the Slovak population, but in many ways paralyzed the normal functioning of our society. The adopted hygienic and epidemiological measures for the prevention of the transmission and spread of the disease, as well as the economic uncertainty and psychological effect of fear of an unknown situation could also negatively affect the decision-making processes of young people in the implementation of some long-term goals. Marriage related decisions can also be included in this group. The aim of the paper is to use the available data from 2020 to analyse especially the intensity and timing of nuptiality of single persons in Slovakia and its possible changes, associated with the specific situation in the pandemic year 2020.

KLÚČOVÉ SLOVÁ

sobášnosť, tabuľky sobášnosti slobodných, intenzita, časovanie, Slovensko, COVID-19, rok 2020

KEY WORDS

nuptiality, gross nuptiality tables, quantum, timing, Slovakia, COVID-19, year 2020

1. ÚVOD

Celosvetová pandémia ochorenia COVID-19 v roku 2020 zasiahla a negatívnym spôsobom ovplyvnila nielen zdravotný stav a úmrtnostné pomery obyvateľstva Slovenska, ale v kombinácii s prijímanými epidemiologickými a čiastočne aj ekonomickými opatreniami (prerušenie výroby, zatváranie prevádzok a pod.) a následne aj samotnými dopadmi nastupujúcej hospodárskej krízy (prepad HDP)

¹ Príspevok je výsledkom projektu VEGA č. 2/0064/20 Pokračujúca transformácia rodinného a reprodukčného správania na Slovensku v časovom a priestorovom aspekte.

vytvorila spoločenskú klímu, v ktorej realizáciu dlhodobých zámerov mohlo vo väčšej miere obyvateľstvo odkladať na príhodnejšie obdobie. Takáto reakcia v podobe strategického odkladania sa môže dotýkať nielen ekonomických aspektov (napr. v podobe zvyšovania úspor, odkladania kúpy nákladnejších tovarov, nehnuteľností a pod.), ale často ju identifikujeme aj v spojitosti s reprodukčným správaním. Najčastejšie sa v tomto smere analyzuje vývoj procesov sobášnosti a plodnosti. Kým v druhom prípade je potrebné prípadné negatívne reakcie v podobe odkladania rodenia detí očakávať s určitým časovým odstupom, keďže koncepcie z pandemického roka 2020 a najmä obdobia najviac zasiahnutého krízovým stavom budú realizované v prevažnej miere až nasledujúci kalendárny rok, pri sobášoch a teda v procese sobášnosti je možné tieto dôsledky identifikovať priamo v príslušný pandemický rok. Ako už ukazovali predbežné a následne aj definitívne údaje o počte uzavretých manželstiev v roku 2020, tento predpoklad sa jednoznačne naplnil. Cieľom predloženého príspevku je preto podrobná analýza procesu sobášnosti slobodných, ktorá zohráva v tomto demografickom procese kľúčovú úlohu. Zameriame sa pritom nielen na samotný rok 2020 a identifikáciu a rozsah zmien v intenzite, časovaní a vekovom charaktere procesu sobášnosti, ale tieto zmeny zasadíme do širšieho vývojového trendu, ktorý bolo možné na Slovensku pozorovať v posledných približne 10 rokoch. Vďaka tomu získame komplexnejší obraz o rozsahu vplyvu pandemického roka 2020 na proces sobášnosti na Slovensku.

2. ZDROJE ÚDAJOV A METODIKA PRÁCE

Hlavný zdroj údajov na analýzu procesu sobášnosti na Slovensku v roku 2020, ako aj v predchádzajúcom období predstavuje vyčerpávajúce zisťovanie, ktoré každoročne realizuje Štatistický úrad Slovenskej republiky (ďalej ŠÚ SR) v rámci hlásení rady Obyv 1-12 (Hlásenie o uzavretí manželstva). Prostredníctvom nich získavame informácie o všetkých uzavretých manželstvách, resp. oboch snúbencoch na území Slovenska. V našej práci využijeme najmä údaje týkajúce sa rodinného stavu, veku (roku narodenia) a miesta bydliska ženícha a nevesty.

Základom analýzy procesu sobášnosti slobodných osôb bude konštrukcia jednovýchodnej tabuľky sobášnosti slobodných priamou metódou v druhom hlavnom súbore udalostí. Tabuľky sobášnosti patria medzi pokročilejšie demografické modely simulujúce tento demografický proces. Ich hlavnou snahou je priblížiť fungovanie procesu sobášnosti slobodných prostredníctvom prevodu reálnej intenzity na báze pravdepodobností na fiktívnu tabuľkovú populáciu. Týmto postupom je výsledný indikátor intenzity sobášnosti slobodných očistený od zmien a rozdielov (napr. pri porovnávaní dvoch vekovo odlišných populácií) vo vekovej štruktúre a od rodinného stavu. V našom prípade budeme vychádzať z metodických konceptov publikovaných v prácach [5, 6].

Na aplikáciu priamej metódy výpočtu pravdepodobností sobáša slobodných v jednovýchodnej tabuľke potrebujeme disponovať v danom roku údajmi o počte sobášov slobodných, o úmrtiach a migračnom salde slobodných osôb podľa veku a roku narodenia. Exponovanou populáciou, teda populáciou vystavenou šanci vstúpiť do manželstva sú slobodné osoby v dokončenom veku $15^2 - 49$ rokov (k 1.1. roka výpočtu).

² Osoby vo veku 15 rokov síce podľa platnej legislatívy nemôžu vstúpiť do manželstva, ale keďže ide o skupinu slobodných k 1.1., teda na začiatku kalendárneho roka, počas neho dovŕšia 16. narodeniny a stanú sa tým exponovanou populáciou na vstup do prvého manželstva.

Priamy výpočet pravdepodobnosti sobáša slobodných v druhom hlavnom súbore udalostí dáva do pomeru sobáše slobodných osôb realizované v dokončenom veku (x) a ($x+1$) v danom kalendárnom roku k príslušnej exponovanej populácii slobodných na začiatku kalendárneho roka vo veku (x). Ak označíme ${}_{t}^{z,s}q_{x'}^{m/\bar{z}}$ ako pravdepodobnosť sobáša slobodného muža alebo ženy v kalendárnom roku (t) vo veku (x) vstúpiť do manželstva, potom pre druhý hlavný súbor udalostí môžeme napísať:

$${}_{t}^{z,s}q_{x'}^{m/\bar{z}} = \frac{{}_{t}^{z,s}S_{x,x+1}^{m/\bar{z}}}{{}_{1.1.t}^{z,s}P_x^{m/\bar{z}} - \frac{{}_{t}^{z,s}D_{x,x+1}^{m/\bar{z}}}{2} + \frac{{}_{t}^{z,s}I_{x,x+1}^{m/\bar{z}}}{2} - \frac{{}_{t}^{z,s}E_{x,x+1}^{m/\bar{z}}}{2}} \quad (1)$$

pričom platí, že:

${}_{t}^{z,s}S_{x,x+1}^{m/\bar{z}}$ je počet sobášov slobodných mužov alebo žien v dokončenom veku (x) a ($x + 1$), z generácie (z) v roku (t),

${}_{t}^{z,s}D_{x,x+1}^{m/\bar{z}}$ je počet úmrtí slobodných mužov alebo žien v dokončenom veku (x) a ($x + 1$), z generácie (z) v roku (t),

${}_{t}^{z,s}I_{x,x+1}^{m/\bar{z}}$ je počet imigrujúcich slobodných mužov alebo žien v dokončenom veku (x) a ($x + 1$), z generácie (z) v roku (t),

${}_{t}^{z,s}E_{x,x+1}^{m/\bar{z}}$ je počet emigrujúcich slobodných mužov alebo žien v dokončenom veku (x) a ($x + 1$), z generácie (z) v roku (t),

${}_{1.1.t}^{z,s}P_x^{m/\bar{z}}$ je počet slobodných mužov alebo žien v dokončenom veku (x) k 1.1. v kalendárnom roku (t).

Takto získané reálne pravdepodobnosti sobášov slobodných sa následne aplikujú na tabuľkovú kohortu slobodných osôb. Na začiatku v presnom veku 16 rokov (l_{16}) ide o tzv. koreň tabuliek sobášnosti. Ten reprezentuje celú tabuľkovú populáciu vystavenú šanci uzavrieť prvé manželstvo. Najčastejšie je z praktických dôvodov zvolený násobok 10 (v našom prípade išlo o 100 000 osôb).

Takto definovaná tabuľková populácia sa následne vystaví pôsobeniu pravdepodobností sobášnosti slobodných v jednotlivých vekových skupinách do konca reprodukčného obdobia (pre vek 16 – 49 rokov). Výsledkom pôsobenia intenzity sobášnosti slobodných je v každom vekovom intervale tabuľkový počet sobášov slobodných, ako súbor osôb, ktoré z tabuľkovej populácie vystúpili v dôsledku uzavretia manželstva.

Ak označíme tabuľkový počet slobodných mužov / žien dožívajúcich sa presného veku (x') v kalendárnom roku (t) ako ${}_{t}^{s}l_{x'}^{m/\bar{z}}$ a tabuľkové sobáše slobodných mužov alebo žien v dokončenom veku (x) a ($x+1$) v kalendárnom roku (t) ako ${}_{t}^{s}d_{x,x+1}^{m/\bar{z}}$, potom platia nasledujúce vzťahy:

$${}_{t}^{s}d_{x,x+1}^{m/\bar{z}} = {}_{t}^{s}q_{x'}^{m/\bar{z}} \cdot {}_{t}^{s}l_{x'}^{m/\bar{z}} \quad (2)$$

$${}_{t}^{s}l_{x'+1}^{m/\bar{z}} = {}_{t}^{s}l_{x'}^{m/\bar{z}} - {}_{t}^{s}d_{x,x+1}^{m/\bar{z}} \quad (3)$$

Výslednou funkciou sobášnych tabuliek je tabuľkový počet slobodných osôb v presnom veku 50 rokov (l_{50}). Indikuje koľko osôb by pri zachovaní intenzity sobášnosti slobodných z daného kalendárneho roka zostalo slobodných, teda by pri známej intenzite sobášnosti nevstúpili medzi 16. a 49. rokom života do manželstva. Keďže sme v našom prípade koreň tabuliek stanovili na 100 000, potom jednoduchou operáciou v podobe $(l_{50})/1000$ získavame tzv. podiel trvalo slobodných osôb. Na hodnotenie procesu sobášnosti sa však využíva aditívna informácia, a teda celkový počet, resp. podiel osôb, ktoré do manželstva vstúpili. Je možné ju v podstate konštruovať pre akýkoľvek vek (16 – 50 rokov), pričom v demografii sa označuje ako tabuľková prvosobášnosť (a vyjadruje sa v %).

Okrem intenzity sobášnosti slobodných budeme pri hlbšej analýze vplyvu pandemického roka 2020 smerovať svoju pozornosť aj na analýzu časovania tohto procesu. Na tento účel využijeme tabuľkový priemerný vek pri prvom sobáši. Ten predstavuje v podstate priemerný počet rokov (v dokončenom veku) od narodenia osoby po uzavretie prvého manželstva. Konštrukcia sa opiera o tabuľkové počty sobášov slobodných v danom kalendárnom roku. Keďže pracujeme v druhom hlavnom súbore udalostí, konštrukciu tabuľkového priemerného veku pri prvom sobáši môžeme vyjadriť nasledujúcim vzťahom:

$$TPVS_t^{m/z} = \frac{\sum_{x=16}^{49} (x+1) \cdot {}_s d_x^{m/z}}{\sum_{x=16}^{49} {}_s d_x^{m/z}} \quad (4)$$

$TPVS_t^{m/z}$ je tabuľkový priemerný vek muža alebo ženy pri prvom sobáši v roku (t),
 ${}_s d_x^{m/z}$ je tabuľkový počet sobášov slobodných mužov alebo žien v roku (t),
 x vek v dokončených rokoch (v rozmedzí 16 až 49 rokov).

Okrem toho využijeme aj niektoré ukazovatele štatistického rozloženia tabuľkových sobášov s vekom. Pôjde o vybrané kvantily – medián, horný a dolný kvartil tabuľkových sobášov slobodných. Z nich následne budeme konštruovať interkvartilové rozpätie tabuľkových sobášov slobodných mužov a žien, ktorého hodnota umožňuje sledovať, ako je nastavená veková koncentrácia celkovo polovice tabuľkovej sobášnosti slobodných, teda v ako širokom vekovom intervale sa realizuje 50 % z intenzity manželských vstupov slobodných osôb. Pri znižovaní hodnôt tohto ukazovateľa môžeme hovoriť o vekovej homogenizácii a tým koncentracii sobášnosti slobodných a v prípade opačného vývoja o vekovej pluralizácii vstupov do prvého manželstva.

Ako už bolo naznačené, uvedené stredné hodnoty sú konštruované z tabuľkových sobášov slobodných, resp. ich vekových relatívnych početností z celkovej tabuľkovej prvosobášnosti. Tie získavame nasledujúcim vzťahom:

$$f_i = \frac{{}_s d_x^{m/z}}{\sum_{x=16}^{49} {}_s d_x^{m/z}} \quad (5)$$

Pre akýkoľvek kvantil potom môžeme odvodiť:

$$Q_k^{(\alpha)} = a_k + h \cdot \frac{k - \sum_{i=1}^{r-1} f_i}{f_{Q_k}^{(\alpha)}} \quad (6), \text{ pričom}$$

a_k je hodnota dolnej hranice intervalu, v ktorom sa nachádza sledovaný kvantil, h je šírka, rozpätie kvantilového intervalu. Keďže pracujeme s tabuľkami sobášnosti s 1-ročnými vekovými skupinami, je možné h zanedbať,

$\frac{k}{\alpha}$ je k -tý kvantil. Pre dolný kvartil 25, pre medián 50 a pre horný kvartil 75,

$\sum_{i=1}^{r-1} f_i$ je kumulatívna relatívna početnosť intervalu bezprostredne predchádzajúceho kvantilového intervalu,

$f_{Q_k}^{(\alpha)}$ je relatívna početnosť kvantilového intervalu.

Okrem kvantilov a interkvartilového rozpätia na analýzu prípadných zmien rozloženia tabuľkových sobášov slobodných mužov a žien v roku 2020 využijeme aj príspevky vybraných vekových intervalov k celkovej tabuľkovej prvosobášnosti. Budeme pritom pracovať s vekom do 25 rokov a 30 a viac rokov. Kým prvý vekový interval signalizuje koncentráciu sobášnosti slobodných v mladom veku, druhý poukazuje na možné posuny do druhej polovice reprodukčného obdobia.

Podiel tabuľkovej sobášnosti slobodných vo veku do 25 rokov potom vyjadríme ako:

$${}_{t}pTS_{do\ 25r.}^{m/\bar{z}} = \frac{\sum_{x=16}^{24} {}_s d_x^{m/\bar{z}}}{\sum_{x=16}^{49} {}_s d_x^{m/\bar{z}}} \cdot 100 \quad (7), \text{ pričom}$$

${}_{t}pTS_{do\ 25r.}^{poh}$ je podiel tabuľkovej sobášnosti slobodných mužov alebo žien vo veku do 25 rokov,

${}_s d_x^{m/\bar{z}}$ je tabuľkový počet sobášov slobodných mužov alebo žien v roku (t).

Podiel sobášnosti slobodných osôb vo veku 30 a viac rokov vyjadruje váhu sobášnosti slobodných v druhej polovici reprodukčného veku na celkovej intenzite sobášnosti slobodných:

$${}_{t}pTS_{30+r.}^{m/\bar{z}} = \frac{\sum_{x=30}^{49} {}_s d_x^{m/\bar{z}}}{\sum_{x=16}^{49} {}_s d_x^{m/\bar{z}}} \cdot 100 \quad (8), \text{ pričom}$$

${}_{t}pTS_{30+r.}^{poh}$ je podiel sobášnosti slobodných mužov alebo žien vo veku 30 a viac rokov,

${}_s d_x^{m/\bar{z}}$ je tabuľkový počet sobášov slobodných mužov alebo žien v roku (t).

3. ROK 2020 A JEHO VPLYVY NA INTENZITU SOBÁŠNOSTI NA SLOVENSKU

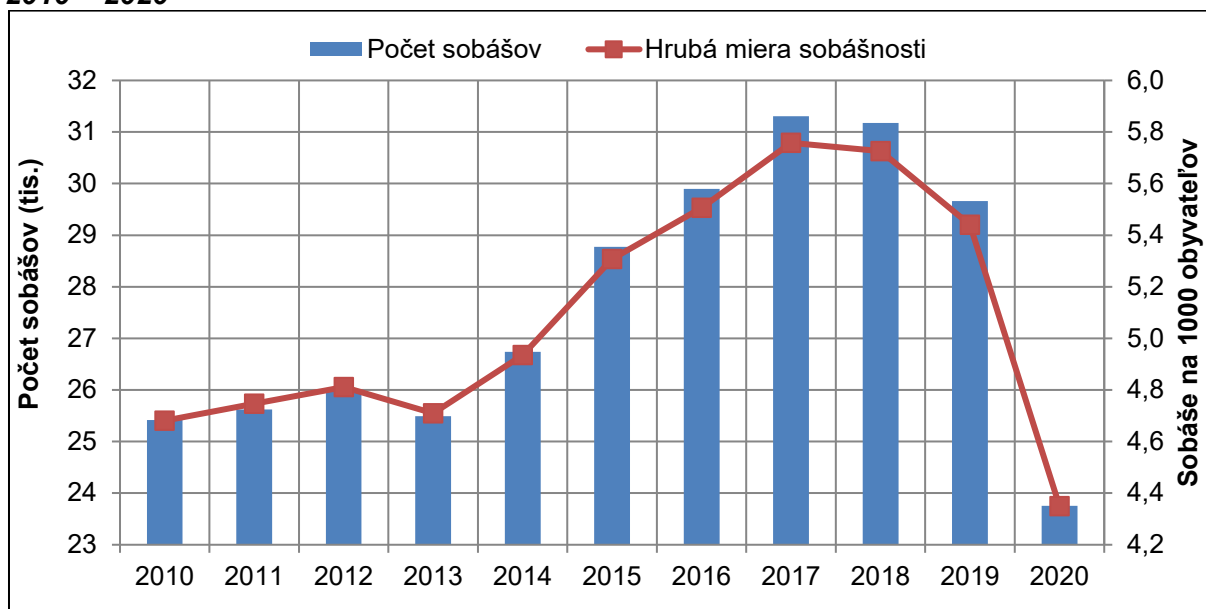
Prvotná a základná informácia, ktorá bola komunikovaná odbornej i laickej verejnosti v spojitosti s dosahom pandemického roku 2020 na sobášnosť na Slovensku, bol počet uzavretých manželstiev. Ak tento prvotný údaj zasadíme do širšieho vývojového kontextu (graf č. 1), môžeme vidieť, že približne od roku 2013 nastalo na Slovensku pomerne významné oživenie počtu sobášov, keď z približne 25 až 26tis. ich počet vzrástol do roku 2017 až nad hranicu 31tis. udalostí. Nad touto hranicou sa udržal aj nasledujúci kalendárny rok, no ďalší medziročný vývoj priniesol

pokles pod úroveň 30tis. Pandemický rok 2020 však v tomto smere bol extrémom, keďže početnosť uzavretých manželstiev sa medziročne znížila pod úroveň 23,8tis. udalostí, čo znamenalo pokles o viac ako 5,9tis. sobášov. Takýto nízky počet sobášov malo Slovensko naposledy na začiatku milénia v roku 2001.

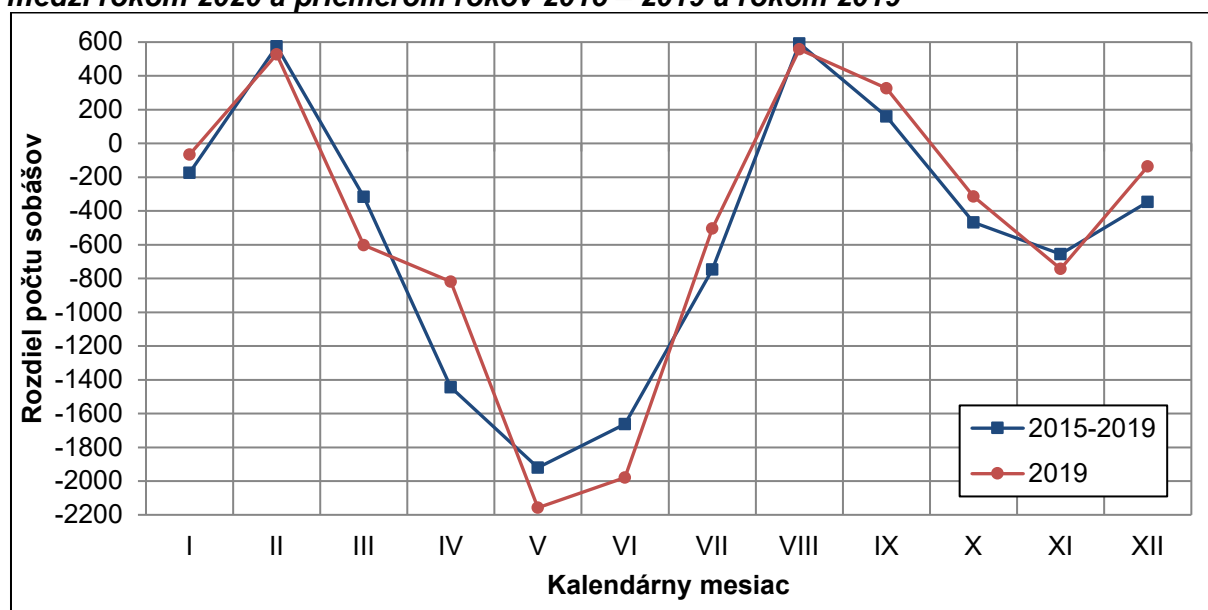
V relatívnom vyjadrení na 1000 obyvateľov pripadalo v rokoch 2010 – 2013 približne 4,6 – 4,8 sobáša. Od roku 2015 sa hrubá miera sobášnosti dostala nad hranicu 5 ‰ a vrchol dosahovala v rokoch 2017 a 2018, keď na 1000 obyvateľov pripadalo viac ako 5,7 sobáša. Spomínaný medziročný pokles počtu sobášov v roku 2019 znížil túto hodnotu na niečo viac ako 5,4 sobáša, no pandemický rok 2020 znamenal bezprecedentný pokles o viac ako jeden promilový bod (na menej ako 4,4 ‰). Takúto nízku hrubú mieru sobášnosti Slovensko dosahovalo naposledy v roku 2001.

Zaujímavosťou tiež je, že spomínaný deficit počtu sobášov v porovnaní s predchádzajúcim obdobím (rokom) významne kopíroval obdobia, keď sa prijímali reštriktívne epidemiologické opatrenia a naopak. Začiatok kalendárneho roka 2020 sa pritom niesol v mierne pozitívnom vývoji, keď najmä vo februári došlo k signifikantnému zvýšeniu počtu uzavretých manželstiev. Môžeme sa len domnievať, že mohlo ísť o určitú reakciu párov na prezentované správy a prijímané opatrenia v niektorých štátoch sveta a obavy z podobného priebehu na Slovensku. Nasledujúce mesiace (od marca do júla) sa však naplno prejavili v podobe výrazného poklesu počtu uzavretých manželstiev, pričom najmä vývoj v období apríl – jún bol kľúčovým pre celkový prepád počtu udalostí na Slovensku. Uvoľňovanie opatrení v letných mesiacoch sa prejavilo len v obmedzenom náraste počtu uzavretých manželstiev, ktorý sa navyše dotýkal len veľmi krátkeho úseku roka (august a september).

Graf č. 1: Vývoj počtu sobášov a hrubej miery sobášnosti na Slovensku v rokoch 2010 – 2020



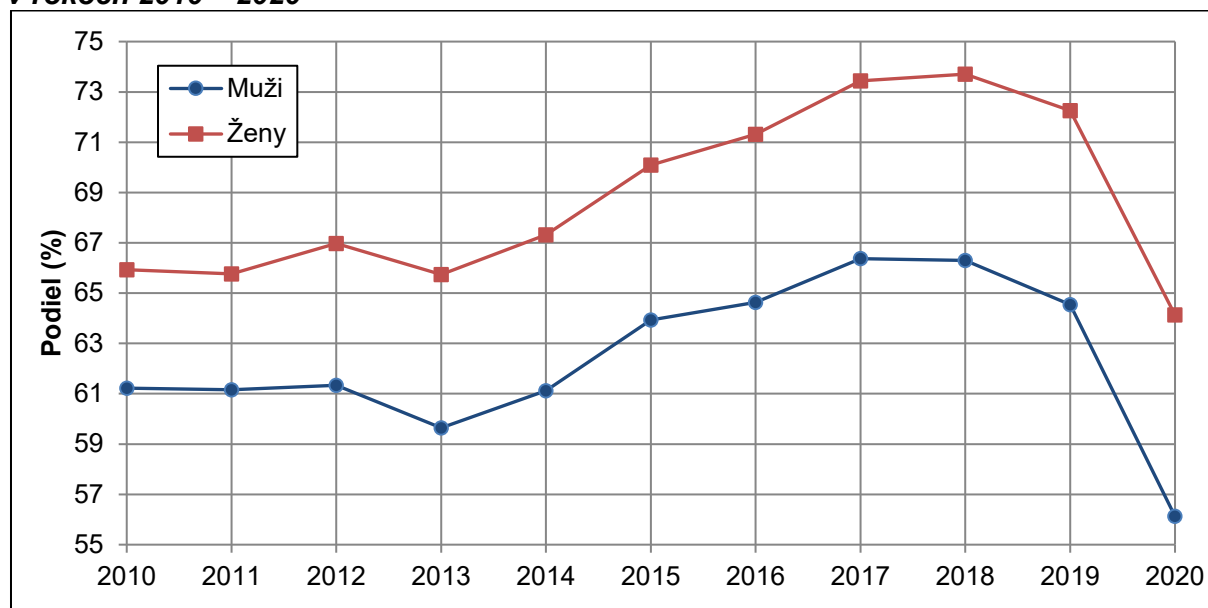
Zdroj údajov: ŠÚ SR DATAcube, výpočty autora

Graf č. 2: Rozdiel počtu uzavretých manželstiev podľa mesiaca sobáša na Slovensku medzi rokom 2020 a priemerom rokov 2015 – 2019 a rokom 2019

Zdroj údajov: ŠÚ SR DATAcube, výpočty autora

Keďže počet sobášov a z nich odvodená hrubá miera sobášnosti predstavujú len základný pohľad do problematiky, ktorý neumožňuje podrobnejšie zhodnotenie vplyvu špecifických podmienok pandemického roka 2020 na proces sobášnosti a jeho jednotlivé aspekty, v ďalšej časti príspevku sa zameriame na informácie, ktoré nám poskytujú niektoré komplexnejšie demografické modely a z nich odvodené indikátory.

Vývoj intenzity sobášnosti slobodných mužov a žien na Slovensku od začiatku tohto storočia zasiahli viaceré medziročné fázy rastu a následného poklesu [1, 7, 8], preto až obdobie po roku 2013 môžeme jednoznačne spájať s dlhším oživením tohto procesu. Ako je zrejme z grafu č. 3, tabuľková prvosobášnosť sa u oboch pohlaví v podstate kontinuálne zvyšovala až do roku 2018. Pri zachovaní intenzity sobášnosti pozorovanej v jednotlivých rokoch by sa podiel slobodných osôb, ktoré aspoň raz do konca reprodukčného veku vstúpili do manželstva, zvýšil u mužov z menej ako 60 % na viac ako 66 % a u žien z menej ako 66 % na takmer 74 %. Medzi rokmi 2018 a 2019 je možné identifikovať určitý pokles, u mužov na úroveň 64,5 % a u žien na niečo viac ako 72 %.

Graf č. 3: Vývoj tabuľkovej prvosobášnosti mužov a žien vo veku 50 rokov na Slovensku v rokoch 2010 – 2020

Zdroj údajov: výpočty autora

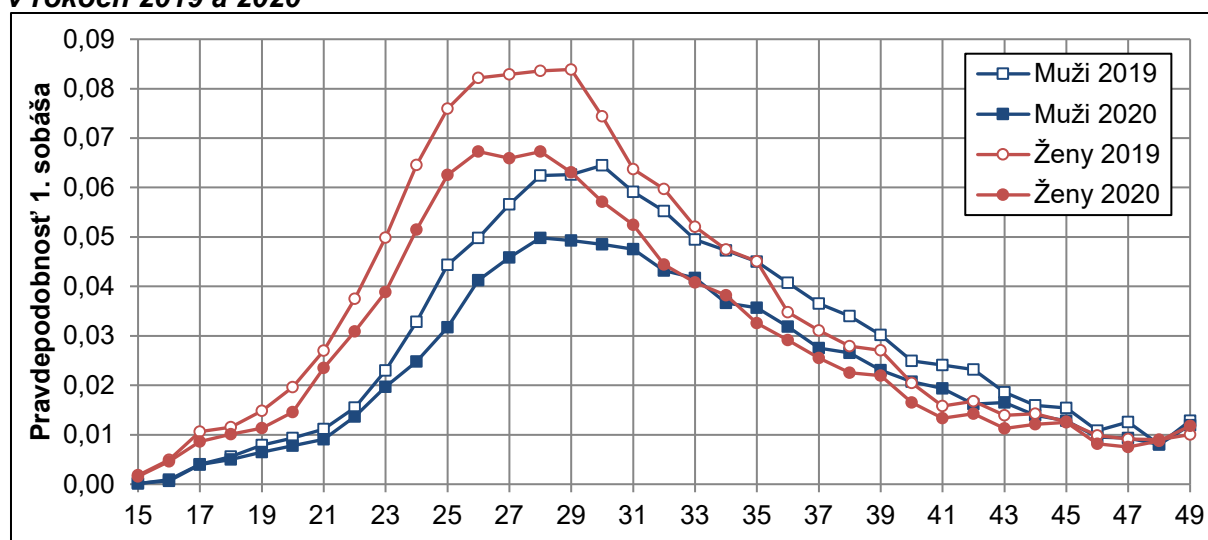
Špecifická situácia, ktorú priniesla pandémia ochorenia COVID-19 v roku 2020 však vyústila do veľmi výrazného poklesu intenzity sobášnosti slobodných. V absolútnom vyjadrení sa tabuľková prvosobášnosť v mužskej časti populácie znížila o 8,4 p. b. a u žien o 8,1 p. b. Znamená to, že pri zachovaní tejto intenzity sobášnosti slobodných by do manželstva vstúpilo len niečo viac ako 56 % mužov a 64 % žien. Takúto nízku sobášnosť slobodných Slovensko vo svojej modernej histórii ešte nezaznamenalo.

Detailné porovnanie pravdepodobností sobáša slobodných podľa veku medzi rokmi 2019 a 2020 (graf č. 4) potvrdilo, že k poklesu šancí vstúpiť do manželstva došlo v pandemickom roku 2020 v podstate vo všetkých vekových skupinách, a to u oboch pohlaví. U žien najvýraznejšie klesla pravdepodobnosť sobáša slobodných vo veku 29 rokov a následne v okolitom veku (27, 28 a 30 rokov). V mužskej časti populácie maximum poklesu nachádzame vo veku 31 rokov, pričom s určitým odstupom nasledovali aj vek 28, 29 a 31 rokov. V dôsledku tohto vývoja tiež nastal určitý posun časovania maximálnej úrovne sobášnosti slobodných osôb. Kým v roku 2019 to u mužov bolo vo veku 30 rokov, v roku 2020 ho identifikujeme už o dva roky skôr (graf č. 4). U žien maximum v prvom analyzovanom roku vznikalo vo veku 29 rokov, kým špecifické pandemické podmienky v nasledujúcom kalendárnom roku prispeli k jeho posunu do veku 26 a 28 rokov. Uvedené „omladnutie“ vekového rozloženia sobášnosti slobodných, ktoré so sebou prinieslo neštandardné obdobie minulého kalendárneho roka, budeme hlbšie analyzovať v nasledujúcej kapitole tohto príspevku. Získané výsledky tak potvrdili, že výrazný prepád počtu sobášov, ako aj intenzity, s akou do prvého manželstva vstupovali muži a ženy na Slovensku v roku 2020, bolo výsledkom poklesu sobášnosti, ktorý sa síce dotýkal všetkých vekových skupín, no pri určitom zovšeobecnení sa dá povedať, že sa najviac dotkol intervalov vyznačujúcich sa v súčasnosti vysokou až najvyššou úrovňou sobášnosti. Aj to by vysvetľovalo, prečo identifikovaný celkový prepád tabuľkovej prvosobášnosti bol medziročne taký razantný.

Pokles pravdepodobností sobášnosti slobodných sa v roku 2020 prejavil nielen na spomenutej celkovej prvosobášnosti, ale jeho dosah môžeme pozorovať aj v ďalších

vybraných vekoch. Na grafoch č. 5 a 6 sú zobrazené tabuľkové podiely slobodných mužov a žien dožívajúcich sa určitého presného veku, teda tabuľková populácia, ktorá sa „vyhla“ uzavretiu manželstva do určitého veku. V prezentovaných údajoch sa po roku 2013 prejavuje pozitívne smerovanie sobášnosti na Slovensku, keď u oboch pohlaví môžeme vidieť postupne medziročné klesanie podielu tabuľkových slobodných osôb, a to v podstate s výnimkou najmladšieho veku vo všetkých vybraných presných vekoch.

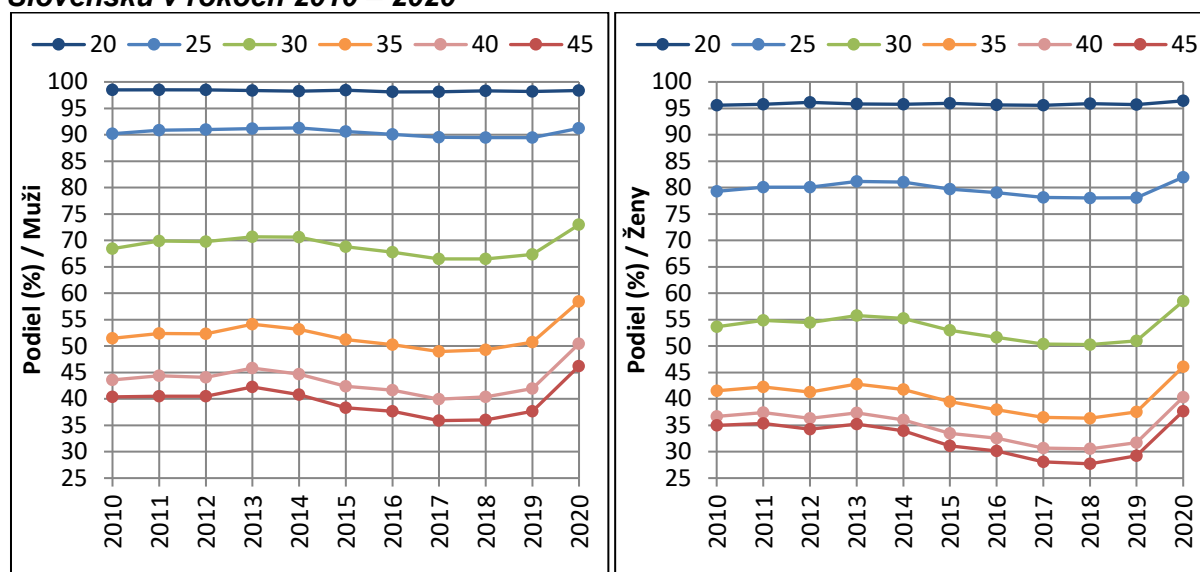
Graf č. 4: Pravdepodobnosť sobáša slobodných mužov a žien na Slovensku podľa veku v rokoch 2019 a 2020



Zdroj údajov: výpočty autora

Keďže medzi rokmi 2018 a 2019 došlo k spomínanému miernemu zníženiu intenzity sobášnosti, prejavilo sa to aj v podieloch tabuľkových slobodných osôb a konkrétne v ich náraste. Dynamika tohto vývoja však bola neporovnateľná s tým, čo znamenal pre sobášnosť slobodných pandemický rok 2020. Identifikovaný výrazný pokles sobášnosti, ktorý sa dotkol v podstate všetkých vekových skupín, sa akumuloval v najväčšej miere logicky v najvyšších vekoch. Napríklad u mužov vo veku 40 a 45 rokov došlo k nárastu podielu tabuľkových slobodných o viac ako 8,5 p. b. V prípade žien to vo veku 35 a 40 rokov bolo dokonca 8,6 p. b.

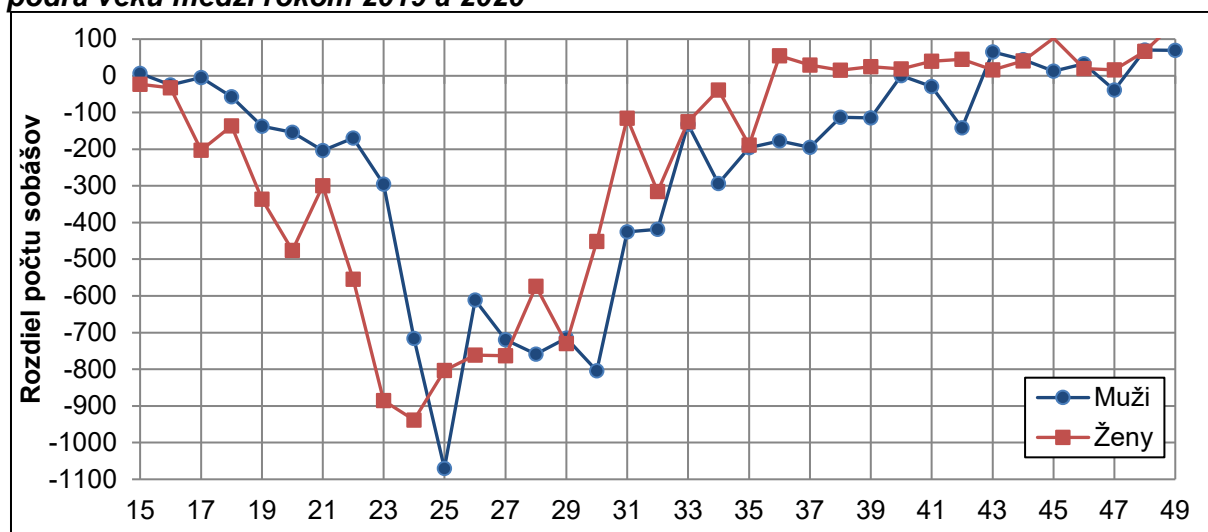
Graf č. 5 a č. 6: Tabuľkové podiely slobodných osôb vo vybraných presných vekoch na Slovensku v rokoch 2010 – 2020



Zdroj údajov: výpočty autora

Asi najvýstižnejšie môžeme identifikovať priame dôsledky špecifickej situácie pandemického roka 2020 na proces sobášnosti slobodných mužov a žien prostredníctvom rozdielu tabuľkového počtu sobášov podľa veku. Ako sme už spomenuli, v mužskej časti populácie sa medzi rokom 2019 a 2020 znížil ich počet o približne 8,4 tis. a u žien to bolo o 8,1tis. Ak sa pozrieme detailnejšie na rozloženie tejto diferencie podľa veku, vidíme, že hlavná časť poklesu celkovej prvosobášnosti sa koncentrovala približne vo veku 24 – 30 rokov. Išlo o takmer 5,4tis. tabuľkových sobášov, teda takmer dve tretiny z poklesu.

Graf č. 7: Rozdiel tabuľkového počtu sobášov slobodných mužov a žien na Slovensku podľa veku medzi rokom 2019 a 2020



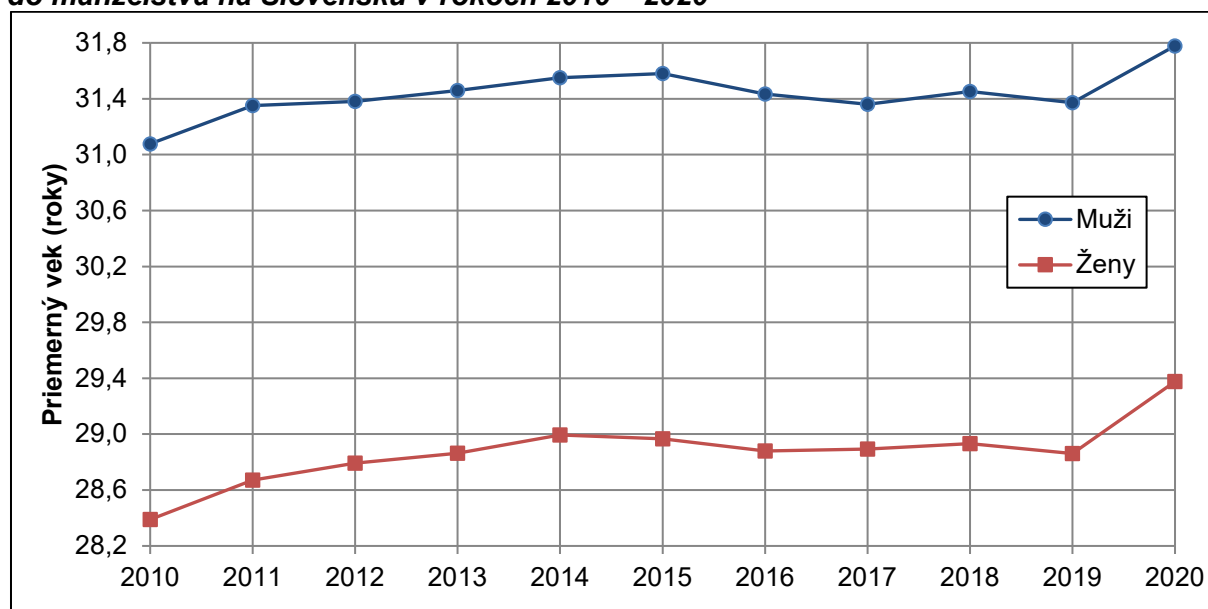
Zdroj údajov: výpočty autora

U žien môžeme identifikovať pre vývoj sobášnosti ako hlavný vek 23 – 29 rokov s poklesom takmer o 5,5tis., čo v relatívnom vyjadrení predstavovalo viac ako dve tretiny z daného medziročného poklesu.

4. ROK 2020 A JEHO VPLYV NA ČASOVANIE SOBÁŠNOSTI SLOBODNÝCH NA SLOVENSKU

Identifikovaný posun v maximálnej úrovni, ako aj rozloženie hlavných vekových skupín prispievajúcich k celkovému poklesu tabuľkovej prvosobášnosti mužov a žien na Slovensku medzi rokom 2019 a 2020 predznamenávajú skutočnosť, že okrem intenzity mali špecifické podmienky pandemického roka 2020 vplyv aj na časovanie sobášnosti. Ako je zrejmé z grafu č. 8, v podstate až do roku 2014, resp. 2015 u mužov prebiehalo kontinuálne, aj keď už značne obmedzené zvyšovanie tabuľkového priemerného veku slobodných mužov a žien pri vstupe do manželstva. Toto obdobie tak nadväzovalo na predchádzajúci pomerne dynamicky prebiehajúci proces odkladania manželského štartu do vyššieho veku, ktorý bol na Slovensku naštartovaný začiatkom 90. rokov [8, 9]. Vďaka tomu sa úplne eliminoval model skorej sobášnosti [bližšie k rozdielom v Európe, napr. 2, 3, 4], ktorý v našom priestore môžeme empiricky pozorovať minimálne počas celého 20. storočia [2, 8, 9]. Podľa údajov z jednovýhodných tabuliek sobášnosti slobodných mužov sa tabuľkový priemerný vek mužov pri vstupe do prvého manželstva dostal v roku 2015 už takmer na úroveň 31,6 roka a u žien svoje maximum dosiahol v roku 2014 na 29 rokoch. Nasledujúce obdobie prinieslo len mierny pokles a následnú stabilizáciu na približne 31,4 roka u mužov a 28,9 roka u žien. Pandemický rok 2020 sa však z tohto vývojového trendu vymyká, keďže v porovnaní s predchádzajúcim nastal výrazný nárast hodnôt tabuľkového priemerného veku slobodných pri uzavretí manželstva. V prípade mužov to bolo o viac ako 0,4 roka a u žien o dokonca viac ako 0,5 roka. U oboch pohlaví tak v roku 2020 dosiahol tabuľkový priemerný vek pri prvom sobáši na Slovensku svoje historické maximum (31,8 roka muži, 29,4 roka ženy).

Graf č. 8: Vývoj tabuľkového priemerného veku slobodných mužov a žien pri vstupe do manželstva na Slovensku v rokoch 2010 – 2020

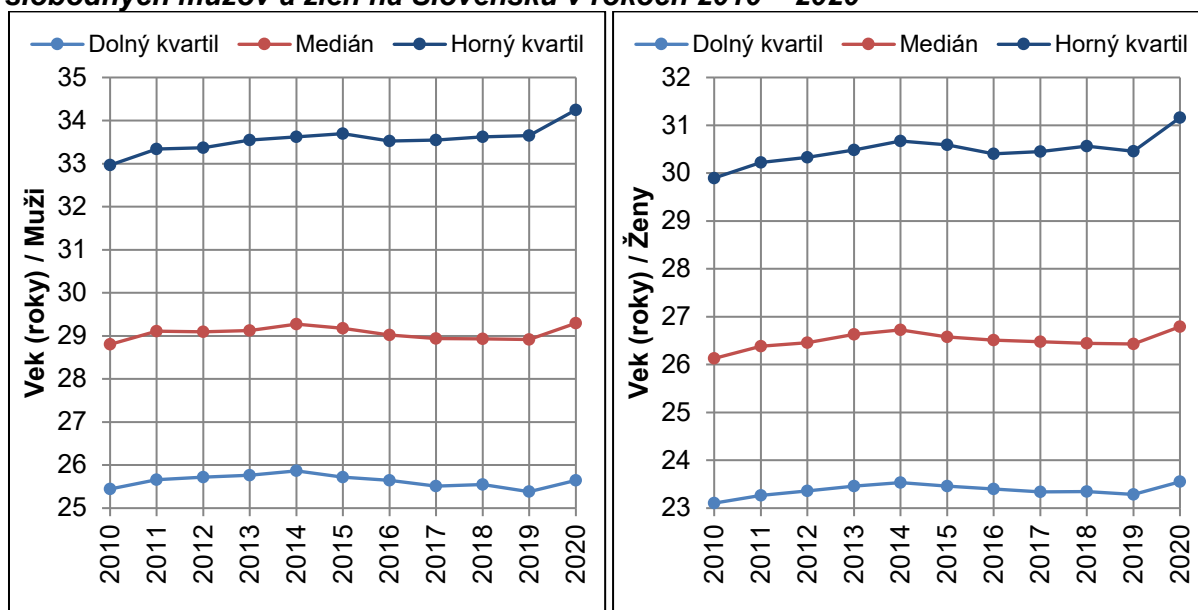


Zdroj údajov: výpočty autora

Rovnako na zmenené podmienky pomerne dynamicky reagovali aj ďalšie ukazovatele časovania sobášnosti slobodných. Ako je zrejmé z grafu č. 9 a 10, u oboch pohlaví medzi rokmi 2019 a 2020 narástli hodnoty dolného i horného kvartilu tabuľkových sobášov, ako aj ich vekového mediánu. Vďaka tomu došlo v porovnaní s predchádzajúcimi rokmi k pomerne významnej vývojovej dynamizácii.

Vzhľadom na tieto skutočnosti, keď zmeny v intenzite sobášnosti nastali najmä vo veku najvyššej sobášnosti, je tiež logickým záver o výraznejšom posune (0,6 roka muži, 0,7 roka ženy) horného kvartilu. Ten u mužov prelomil prvýkrát hranicu 34 rokov a u žien sa dostal nad úroveň 31 rokov. Rovnako mierne vzrástol aj dolný kvartil (o niečo viac ako štvrt' roka u oboch pohlaví), no svojou hodnotou sa až tak výrazne neodlišoval od úrovni, ktoré tento indikátor dosahoval v poslednej dekáde (graf č. 9 a č. 10). Podľa hodnoty vekového mediánu tabuľkových sobášov by sa pri zachovaní ich vekového rozloženia z roku 2020 polovica prvých sobášov mužov realizovala do veku 29,3 roka a u žien do veku 26,8 roka. V porovnaní s predchádzajúcim rokom 2019 tak došlo k nárastu o približne 0,38 roka u mužov, resp. 0,36 roka u žien.

Graf č. 9 a č. 10: Dolný, horný kvartil a vekový medián tabuľkovej sobášnosti slobodných mužov a žien na Slovensku v rokoch 2010 – 2020



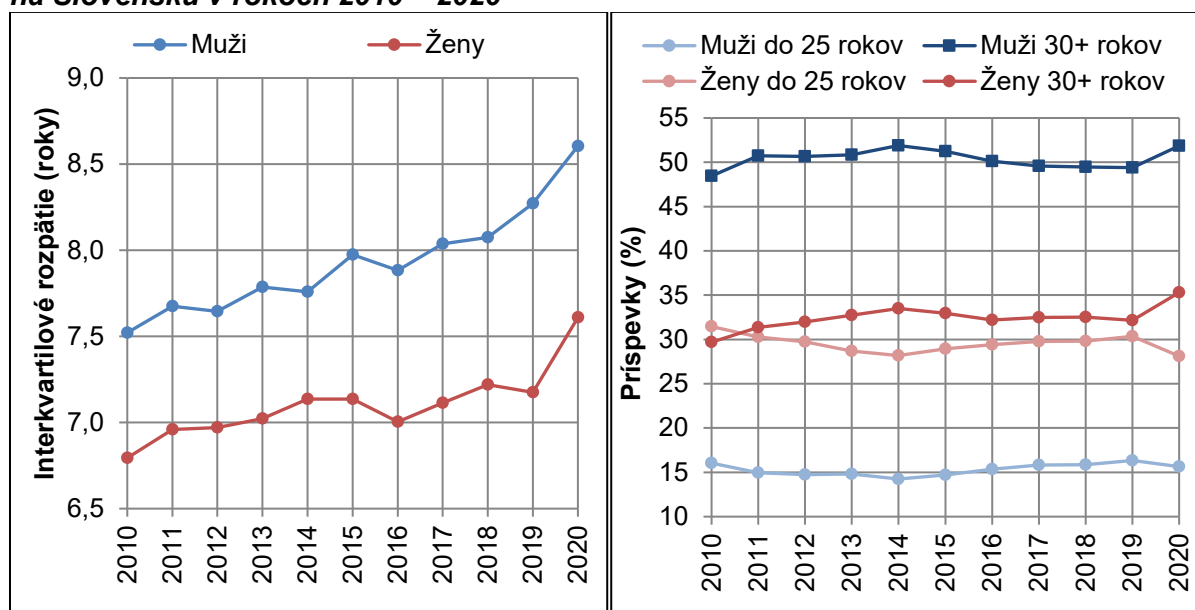
Zdroj údajov: výpočty autora

Keďže hodnota horného kvartilu vzrástla rýchlejšie, ako tomu bolo v prípade dolného kvartilu, špecifické podmienky roka 2020 priniesli aj pomerne významné predĺženie interkvartilového rozpätia. Aj keď sme dlhodobo svedkami zvyšovania jeho hodnoty, posledná medziročná dynamika ďaleko presiahla doterajšie zmeny (graf č. 11). Kým na začiatku druhej dekády tohto milénia sa u mužov polovica celej tabuľkovej sobášnosti slobodných koncentrovala do vekového intervalu so šírkou niečo viac ako 7,5 roka a do roku 2019 sa predĺžila len na necelých 8,3 roka, v pandemickom roku to už bolo približne 8,6 roka. Obdobne u žien v roku 2010 interkvartilové rozpätie dosahovalo hodnotu necelých 6,8 roka, do roku 2019 sa dostalo na necelých 7,2 roka, aby následne o rok neskôr došlo k prudkému nárastu na viac ako 7,6 roka.

Zmeny v intenzite, časovaní a samotnom rozložení sobášnosti slobodných mužov a žien na Slovensku spojené so špecifickými podmienkami pandemického roka 2020 sa odzrkadlili aj na príspevkoch vybraných vekových skupín k celkovej tabuľkovej sobášnosti (graf č. 12). Predovšetkým sme svedkami signifikantného medziročného nárastu podielu tabuľkovej sobášnosti realizovanej vo veku 30 a viac rokov. U mužov sa ich zastúpenie medzi rokmi 2019 a 2020 zvýšilo z približne 49,4 % na takmer 52 %, pričom posledný vývoj sa skôr niesol v znamení opačného trendu. Rovnako to platilo aj v ženskej časti populácie. V jej prípade však bol daný nárast ešte dynamickejší,

keďže váha tabuľkovej sobášnosti z druhej polovice reprodukčného obdobia vzrástla z niečo viac ako 32 % nad hranicu 35 %. Opačný bol vývoj príspevkov vo veku do 25 rokov. U oboch pohlaví teda identifikujeme pokles, ktorý opätovne prebiehal o niečo dynamickejšie u žien. Znamená to tiež, že špecifické podmienky pandemického roka 2020 zatiaľ zastavili nastúpený trend mierneho posilňovania realizovanej tabuľkovej sobášnosti slobodných mužov a žien v mladom veku (pozri graf č. 12).

Graf č. 11 a č. 12: Interkvartilové rozpätie tabuľkových sobášov slobodných a príspevky vybraných vekových skupín k celkovej tabuľkovej sobášnosti slobodných mužov a žien na Slovensku v rokoch 2010 – 2020



Zdroj údajov: výpočty autora

5. ZÁVER

Špecifické podmienky, ktoré priniesla pre spoločnosť na Slovensku pandémia ochorenia COVID-19 v roku 2020 výrazným spôsobom zasiahli nielen zdravotný stav obyvateľstva, a niektoré výkonové ukazovatele národného hospodárstva, ale pomerne promptne reagoval aj proces sobášnosti. Potvrdilo sa, že sobášnosť na Slovensku predstavuje demografický proces, ktorý je veľmi citlivý na externé podmienky. Výrazné obmedzenia pohybu, prijímané v snahe zabrániť šíreniu ochorenia, mali následne pomerne rozsiahly dosah na uzatváranie manželstiev. Výsledky našej analýzy založené na údajoch tabuliek sobášnosti slobodných potvrdili signifikantný medziročný prepád tabuľkovej prvosobášnosti, keď pravdepodobnosť uzavrieť manželstvo klesla v podstate vo všetkých vekových skupinách. Najvýraznejšie sa to však dotklo vekov, ktoré sa vyznačovali vyššou intenzitou sobášnosti slobodných. S tým boli spojené aj určité zmeny vo vekovom rozložení tabuľkových sobášov. Zaujímavosťou pritom bolo, že vo vyššom veku a na konci reprodukčného obdobia nedošlo v podstate k žiadnemu poklesu počtu tabuľkových sobášov a dokonca sme boli svedkami určitého mierneho medziročného zvýšenia.

Aj vďaka tomu tak v roku 2020 pomerne významne narástol priemerný vek pri prvom sobáši u oboch pohlaví, rovnako sa zvyšovala tiež hodnota horného a dolného kvantilu, či vekového mediánu. Z pohľadu interkvartilového rozpätia je zrejmé, že špecifické podmienky pandemického roka 2020 prispeli k urýchleniu vekovej heterogenizácie uzatvárania prvých sobášov mužmi i ženami na Slovensku. Súčasne

s tým môžeme identifikovať výrazný medziročný nárast príspevkov druhej polovice reprodukčného veku k celkovej prvosobášnosti a naopak, určitý pokles v mladom veku do 25 rokov. V oboch prípadoch bol tento proces o niečo dynamickejší v ženskej časti populácie.

Analyzovaný vplyv pandemického roka 2020 na proces sobášnosti predstavuje ukážku výrazných prierezových zmien, ktoré na celkovú realizáciu manželských zväzkov nemusia mať v podstate žiadny efekt. Na druhej strane sa však predsa len môže vynárať obava, či aj napriek tomu časť párov, ktorým nepriaznivá situácia zmarila snahu o uzavretie manželstva, pri jeho odložení na príhodnejší čas neprehodnotí svoj úmysel. Určité signály v tomto smere by už mohol priniesť vývoj v roku 2021, resp. pokrízové obdobie a s ním spojený predpoklad o formovaní tzv. kompenzačnej fázy.

LITERATÚRA

- [1] FIALA, T. – LANGHAMROVÁ, J. – PECHHOLDOVÁ, M. – ĎURČEK, P. – ŠPROCHA, B.: Population Development of Czechia and Slovakia after 1989. In: Demografie, 2018, č. 3, s. 202 – 218.
- [2] FIALOVÁ, L.: Vývoj sňatečnosti v Československu v letech 1918–1988. In: Historická demografie, 1992, roč. 16, s. 116 – 134.
- [3] HAJNAL, J.: European Marriage Patterns in Perspective. In: Glass, D. V. – D. E. C. Eversley (eds.): Population in History. Essays in Historical Demography. New Brunswick: Aldine Transaction, 1965, s. 101 – 143.
- [4] MONNIER, A. – RYCHTAŘÍKOVÁ, J.: The Division of Europe into East and West. In: Population: An English Selection, 1992, 4, s. 129 – 160.
- [5] PAVLÍK, Z. – RYCHTAŘÍKOVÁ, J. – ŠUBRTOVÁ, A.: Základy demografie. Praha: Academia. 198. 732 s.
- [6] RYCHTAŘÍKOVÁ, J.: Tabulky sňatečnosti a metody jejich konstrukce. In: Demografie, 1984, č. 2, s. 110 – 121.
- [7] ŠPROCHA, B.: Transformácia sobášnosti slobodných na Slovensku optikou tabuliek života. In: Slovenská štatistika a demografia, 2020, č. 4, s. 13 – 27.
- [8] ŠPROCHA, B. – TIŠLIAR, P.: 100 rokov obyvateľstva Slovenska. Od vzniku Československa po súčasnosť. Bratislava: Centrum pre historickú demografiu a populačný vývoj Slovenska. Filozofická fakulta UK Bratislava, 2018.
- [9] VAŇO, B. et al.: Obyvateľstvo Slovenska 1945–2000. Bratislava: INFOSTAT, 2001. 74 s.

RESUMÉ

Špecifické podmienky pandemického roka 2020 ovplyvnili chod celej spoločnosti na Slovensku. V oblasti demografického výskumu sa najväčšia pozornosť venovala dosahu na zdravotný stav obyvateľstva a proces úmrtnosti. Ako však už predznamenal vývoj počtu sobášov, určité prejavy bolo možné očakávať aj v procese uzatvárania manželských zväzkov. Hlavným cieľom príspevku bola analýza sobášnosti slobodných osôb na Slovensku v roku 2020 (v porovnaní s vývojom v predchádzajúcej dekáde) prostredníctvom tabuliek sobášnosti. Svoju pozornosť sme upriamili na zmeny v intenzite, časovaní a vekovom rozložení tabuľkovej sobášnosti slobodných mužov a žien. Získané výsledky potvrdili výrazný pokles pravdepodobnosti prvého sobáša v podstate vo všetkých vekových skupinách. Dôsledkom tohto vývoja bol aj značný medziročný prepád tabuľkovej prvosobášnosti vo veku 50 rokov.

Podľa údajov z roku 2020 by pri zachovaní takejto intenzity sobášnosti slobodných do manželstva vstúpilo len niečo viac ako 56 % mužov a 64 % žien. Takúto nízku sobášnosť slobodných Slovensko vo svojej modernej histórii ešte nezaznamenalo.

Značnými zmenami prešlo aj časovanie sobášnosti. Sme svedkami výrazného nárastu priemerného veku pri prvom sobáši, vekového mediánu, ako aj dolného a horného kvartilu tabuľkovej sobášnosti. Vzrástla tiež veková heterogenita realizácie tabuľkových sobášov, keď polovica z nich bola realizovaná v širšom interkvartilovom rozpätí.

Uvedené zistenia tak potvrdili, že sobášnosť na Slovensku promptne a pomerne prudko reagovala na zmenené podmienky a uvedené zmeny sa dotkli v podstate všetkých dôležitých aspektov tohto demografického procesu. Súčasne s tým je však potrebné podotknúť, že tieto zmeny sú s najväčšou pravdepodobnosťou len dočasné ako reakcia na reštrikčné opatrenia týkajúce sa špecifickej epidemickej situácie na Slovensku a s ňou spojené opatrenia.

RESUME

The specific conditions of the pandemic year 2020 affected the functioning of the whole society in Slovakia. In the field of demographic research, the greatest attention was paid to the impact on the health status of the population and the mortality process. However, as the development of the number of marriages already heralds, certain manifestations could also be expected in nuptiality. The main goal of the paper was to analyse the marriage process of single persons in Slovakia in 2020 (compared to the development in the previous decade) using gross nuptiality tables. We focused our attention on changes in the intensity, timing and age distribution of first table marriages of men and women. The obtained results confirmed a significant decrease in the probability of the first marriage in virtually all age groups. The result of this development was also a significant year-on-year drop in the table first marriage rate at the age of 50.

According to 2020 data, with maintaining such intensity of first marriage rates, only more than 56% of men and 64% of women would enter into marriage. In its modern history, Slovakia has not yet experienced such a low first marriage rate of single persons.

The timing of nuptiality has also undergone significant changes. We are witnessing a significant increase in the mean age at first marriage, the median age, as well as the lower and upper quartiles of table first marriages. The age heterogeneity of table first marriages also increased, with half of them being carried out in a wider interquartile range.

The above findings thus confirmed that nuptiality in Slovakia reacted promptly and relatively sharply to the different conditions, and the afore-mentioned changes affected basically all important aspects of this demographic process. However, it is important to note that these changes are most likely only temporary in response to the restrictive regulations concerning the specific epidemiological situation in Slovakia and the measures associated with it.

PROFESIJNÝ ŽIVOTOPIS

RNDr. Branislav Šprocha, PhD., absolvoval magisterské štúdium na Prírodovedeckej fakulte Univerzity Karlovej v Prahe v odbore demografia a geodemografia (2006). V roku 2011 ukončil doktorandské štúdium v programe demografia a v roku 2021 sa na Prírodovedeckej fakulte Univerzity Komenského v Bratislave habilitoval v odbore Humánna geografia. Od roku 2007 je vedeckovýskumným pracovníkom Výskumného demografického centra pri INFOSTATE a od roku 2009 vedeckým pracovníkom Prognostického ústavu Centra spoločenských

a psychologických vied SAV. V roku 2015 sa stal vedúcim Výskumného demografického centra. V oblasti demografie sa špecializuje na problematiku rodinného a reprodukčného správania a ich vplyvov na spoločnosť. Okrem toho sa zameriava na analýzu vybraných populačných štruktúr, reprodukčného správania rómskeho obyvateľstva na Slovensku a otázky konštrukcie populačných prognóz.

KONTAKT

branislav.sprocha@gmail.com

Mikuláš CĀR

Slovenská štatistická a demografická spoločnosť

HĽADANIE ÚČINNÝCH METÓD A NÁSTROJOV NA HODNOTENIE EPIDEMICKEJ SITUÁCIE

SEARCHING FOR EFFECTIVE METHODS AND TOOLS TO ASSESS AN EPIDEMIOLOGICAL SITUATION

ABSTRAKT

Epidemická situácia v dôsledku vzniku a šírenia sa infekčného ochorenia covid-19 má v doterajšom priebehu cyklický charakter. Na začiatku roka 2022 sme evidovali už štvrtú vlnu pandémie koronavírusu a je veľmi pravdepodobné, že budeme svedkami aj ďalších vln. Efektívne manažovanie boja proti koronavírusu prispieva významnou mierou ku relatívne primeranému fungovaniu slovenskej spoločnosti aj v čase pandémie. Jeho dôležitou súčasťou je aj výber a použitie vhodných ukazovateľov na sledovanie a hodnotenie epidemickej situácie.

ABSTRACT

The epidemiological situation due to the emergence and spread of the Covid-19 pandemic has so far been cyclical. At the beginning of 2022, the fourth wave of the pandemic was recorded and we will possibly see other waves as well. The effective management of the fight against coronavirus contributes significantly to the relatively adequate functioning of Slovak society even in times of pandemic. Its important part of management is also the selection and use of appropriate indicators to monitor and evaluate the epidemiological situation.

KLÚČOVÉ SLOVÁ

vhodný štatistický ukazovateľ, ukazovateľ vážené reprodukčné číslo, hodnotenie vývoja pandémie koronavírusu

KEY WORDS

appropriate statistical indicator, weighted reproduction number indicator, evaluation of the development of a coronavirus pandemic

1. ÚVOD

V priebehu posledných dvoch rokov je náš každodenný život veľmi silno determinovaný epidemickou situáciou v dôsledku šírenia sa infekčného ochorenia covid-19, ktoré je vyvolané koronavírusom SARS-CoV-2. Doterajší vývoj ochorenia covid-19 má cyklický charakter. Na začiatku roka 2021 sme evidovali vrchol druhej vlny, na začiatku decembra 2021 vrchol tretej vlny a v polovici roka 2022 vrchol už štvrtej vlny šírenia sa tejto pandémie. Podľa názoru odborníkov je veľmi pravdepodobné, že v nasledujúcom období budeme čeliť aj ďalším vlnám šírenia koronavírusu. Preto je nevyhnutné stále hľadať účinné metódy a nástroje na reálne hodnotenie epidemickej situácie a aj na jej efektívne manažovanie.

Na základe doterajších skúseností z hodnotenia vývoja epidemickej situácie možno zovšeobecniť určité zaujímavé empirické poznatky, ktoré môžu byť užitočné pri jej hodnotení v nasledujúcich mesiacoch. V našom príspevku poukážeme na určité

nedostatky ukazovateľa počet novo pozitívne testovaných osôb pri hodnotení rizika šírenia koronavírusu na celoštátnej aj regionálnej úrovni, zamyslíme sa nad vylúčením predtým veľmi frekventovaného ukazovateľa reprodukčné číslo zo súboru oficiálnych ukazovateľov na hodnotenie vývoja epidemickej situácie a porovnáme ostatné vlny pandémie koronavírusu pomocou vybraných ukazovateľov.

Efektívne hodnotenie a manažovanie epidemickej situácie predpokladá, aby sa na tento účel používal systém vhodných a hlavne navzájom konzistentných ukazovateľov. Je potrebné dbať na to aby jednotlivé ukazovatele, používané pri sledovaní a hodnotení epidemickej situácie, nevyvolávali pochybnosti o tom, ako reflektujú realitu. Na tieto problémy reaguje druhá a tretia kapitola príspevku. Doterajšie dvojročné skúsenosti sú, bohužiaľ, poznačené aj skúsenosťami z bežného používania nie celkom vhodných ukazovateľov, ale aj nedoceňovaním vzácneho informačného obsahu niektorých ukazovateľov. Vo štvrtej časti príspevku sú použité vybrané ukazovatele, ktoré pomerne dobre odrážajú vývoj epidemickej situácie.

Našu pozornosť zameriame na hodnotenie epidemickej situácie pomocou súboru vybraných ukazovateľov za obdobie od začiatku septembra 2020 do konca februára 2022. Ide v podstate o obdobie, v ktorom sa už plne prejavili druhá až štvrtá vlna pandémie koronavírusu. Začiatky a vrcholy daných vln sa dajú dosť jednoznačne identifikovať na základe hodnôt ukazovateľa miera pozitívne testovaných osôb.

2. DISKUTABILNÉ POUŽÍVANIE UKAZOVATEĽA POČET NOVO POZITÍVNE TESTOVANÝCH OSÔB

Ukazovateľ počet novo pozitívne testovaných osôb patrí nepochybne k tým ukazovateľom, s ktorým sa za posledné dva roky stretávame denno-denne. Je nesporne dôležitým východiskovým ukazovateľom, za určitých predpokladov má aj vypovedaciu schopnosť, ale zároveň patrí k jednému z tých oficiálne používaných ukazovateľov, ku ktorému máme zásadnú vecnú výhradu. Podstata našej výhrady spočíva v tom, že samostatné použitie tohto ukazovateľa prináša často nepresnú a v mnohých prípadoch až zavádzajúcu informáciu. Poukážeme na to na konkrétnom príklade.

To, že ukazovateľ počet novo pozitívne testovaných osôb môže poskytnúť pri priebežnom hodnotení epidemickej situácie aj pri jej komplexnejšom porovnávaní až zavádzajúcu informáciu je spôsobené značnou asymetriou medzi počtom novo pozitívne testovaných osôb a počtom vykonaných PCR testov.

Vôbec neplatí, že pri porovnávaní rizika vývoja koronavírusu vo dvoch časových úsekoch je vyššie riziko automaticky vtedy, keď sa zistí vyšší počet novo pozitívne testovaných osôb. Na vyhodnotenie miery rizikovosti epidemickej situácie v určitom okamihu alebo časovom úseku nie je rozhodujúce, aký počet novo pozitívne testovaných osôb sa odhalí, ale predovšetkým to, z akého celkového počtu zrealizovaných PCR testov vyplynulo dané zistenie.

Rozhodne je vyššia miera rizika šírenia koronavírusu napríklad pri zistení 100 novo pozitívne testovaných osôb prostredníctvom vykonania 500 PCR testov, ako pri zistení 150 pozitívne testovaných osôb prostredníctvom vykonania 750 PCR testov. V prvom prípade pripadá 20 novo pozitívne testovaných osôb na 100 vykonaných PCR testov, kým v druhom prípade len 15 novo pozitívne testovaných na 100 vykonaných PCR

testov. V uvedenom príklade vyšší počet novo pozitívne testovaných reálne automaticky nepredstavuje aj vyššie riziko šírenia koronavírusu práve v dôsledku diskutovanej asymetrie východiskových ukazovateľov.

Ako príklad na výhradu používať absolútny ukazovateľ počet novo pozitívne testovaných osôb na hodnotenie stavu a vývoja epidemickej situácie môže poslúžiť stále pravidelne denne publikovaný prehľad o **počte pozitívne testovaných PCR testami za kraje** na webovej stránke <https://korona.gov.sk/koronavirus-na-slovensku-v-cislach/>.

Na sklonku vrcholiacej tretej vlny pandémie koronavírusu, v prvej polovici decembra 2021, sa na spomenutej stránke uvádzal najvyšší počet novo pozitívne testovaných osôb aj za Bratislavský kraj. Pri detailnejšom pohľade na údaje v náhodne vybraných dňoch s týždňovou frekvenciou v tabuľke č. 1 zistíme, že najvyšší počet novo pozitívne testovaných v danom kraji neznamena automaticky aj najkritickejšiu epidemickú situáciu.

V dňoch 9. 12. 2021 a 16. 12. 2021 bol v Bratislavskom kraji zaznamenaný skutočne najvyšší počet novo pozitívne testovaných osôb spomedzi všetkých slovenských krajov. Napriek tomu bol práve Bratislavský kraj v týchto dňoch najmenej rizikový, lebo mal jednoznačne najnižšiu mieru pozitívne testovaných osôb v porovnaní s ostatnými kraji. To, že počet novo pozitívne testovaných osôb poskytuje v takomto podaní zavádzajúcu informáciu, je spôsobené už uvedenou značnou asymetriou medzi počtom novo pozitívne testovaných a počtom vykonaných PCR testov v daných časových okamihoch v jednotlivých slovenských krajoch.¹

Z údajov v tabuľke č. 1 vyplýva, že v Bratislavskom kraji sa vo všetkých náhodne vybraných decembrových dňoch vykonal najvyšší počet PCR testov spomedzi všetkých slovenských krajov. Môže to súvisieť aj s relatívne najdostupnejšou zdravotnou starostlivosťou v tomto slovenskom regióne.

Tabuľka č. 1: Vzájomne súvisiace ukazovatele za vybrané dni podľa krajov

	BL	TA	TC	NI	ZI	BC	PV	KI
2.12.2021_PT	2 024	1 809	1 793	1 323	1 778	1 756	2 333	1 927
2.12.2021_PCR	7 404	4 956	4 165	3 335	3 943	3 729	6 510	5 101
2.12.2021_MPT	27,3	36,5	43,0	39,7	45,1	47,1	35,8	37,8
9.12.2021_PT	1 017	937	859	675	830	691	693	633
9.12.2021_PCR	3 538	2 468	2 042	1 793	2 069	1 825	2 155	1 660
9.12.2021_MPT	28,7	38,0	42,1	37,6	40,1	37,9	32,2	38,1
16.12.2021_PT	669	656	589	421	645	527	428	351
16.12.2021_PCR	2 808	1 957	1 585	1 295	1 813	1 532	1 645	1 206
16.12.2021_MPT	23,8	33,5	37,2	32,5	35,6	34,4	26,0	29,1

¹ Problém značnej asymetrie medzi počtom novo pozitívne testovaných a počtom vykonaných PCR testov v daných časových okamihoch v jednotlivých slovenských krajoch sa vyskytoval tak pred decembrom 2021, ale aj v januári 2022. Napr. aj 13.1.2022 bol v Bratislavskom kraji zaregistrovaný najvyšší prírastok novo pozitívne testovaných osôb (382 osôb), avšak miera novo pozitívne testovaných osôb v ten deň bola v tomto kraji, podľa podrobnejších okresných údajov, najnižšia zo všetkých slovenských krajov (21,2 %). Trenčiansky kraj s najnižším počtom novo pozitívne testovaných bol však tretí najhorší podľa miery pozitívne testovaných osôb s 27,2 percentami po Žilinskom kraji (30,6 %) a Trnavskom kraji (28,2 %)

23.12.2021_PT	627	557	509	426	711	453	650	437
23.12.2021_PCR	3 861	1 839	1 427	1 359	1 854	1 560	2 197	1 436
23.12.2021_MPT	16,2	30,3	35,7	31,3	38,3	29,0	29,6	30,4
30.12.2021_PT	429	402	373	294	470	327	249	317
30.12.2021_PCR	2 358	1 396	1 084	927	1 341	1 091	1 128	909
30.12.2021_MPT	18,2	28,8	34,4	31,7	35,0	30,0	22,1	34,9

Zdroj: [https://github.com/Institut-Zdravotnych-Analyz/covid19-](https://github.com/Institut-Zdravotnych-Analyz/covid19-data/tree/main/PCR_Tests)

[data/tree/main/PCR_Tests](https://github.com/Institut-Zdravotnych-Analyz/covid19-data/tree/main/PCR_Tests), dátum dostupnosti 30. 12. 2021, vlastné spracovanie

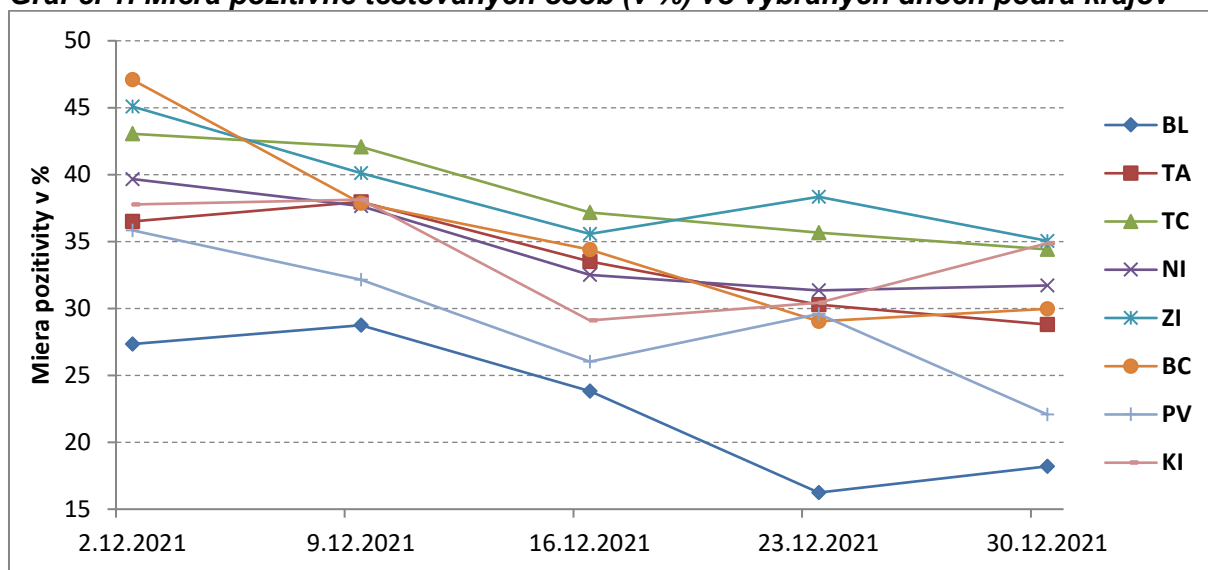
Poznámka: Skratka za dátumom **PT** znamená počet pozitívne testovaných osôb, **PCR** počet vykonaných PCR testov a **MPT** mieru pozitívne testovaných osôb v %.

Zaujímavé je, že v Bratislavskom kraji sa v priebehu decembra 2021 vykonala až viac ako pätina PCR testov v rámci celého Slovenska, pričom je v ňom prihlásených na trvalý pobyt len okolo 12 percent obyvateľov Slovenska. Väčší podiel PCR testov, ako je podiel obyvateľov v rámci Slovenska, sa v priebehu decembra 2021 vykonal ešte aj v Trnavskom kraji. Najväčšia negatívna disproporcía medzi skutočným podielom počtu obyvateľov v rámci Slovenska (takmer 15 %) a podielom počtu vykonaných PCR testov (len necelých 10 %) bola v priebehu decembra 2021 zaznamenaná v Košickom kraji. Uvedené disproporcie sú do značnej miery spôsobené migráciou obyvateľov medzi kraji.

Určitá korekcia uvedenej asymetrie medzi počtom novo pozitívne testovaných osôb a počtom vykonaných PCR testov sa dosiahne napríklad prostredníctvom týždňových priemerov hodnôt týchto dvoch ukazovateľov. Získajú sa tak ich vyrovnanejšie hodnoty, čo dáva daným ukazovateľom vyššiu odbornú štatistickú legitimitu, avšak ani týždňové priemery neriešia vecný problém korektného hodnotenia epidemickej situácie. Korektné vecné riešenie spočíva v použití pomerového ukazovateľa miera pozitívne testovaných osôb, ktorý predstavuje pomer medzi počtom novo pozitívne testovaných osôb a počtom vykonaných PCR testov, zvyčajne vyjadrený v percentách.

Už spomenutú asymetriu medzi počtom novo pozitívne testovaných a počtom vykonaných PCR testov v priebehu náhodne vybraných dní v jednotlivých slovenských krajoch, ktorá je zrejmá z hodnôt v tabuľke č. 1, názorne demonštruje aj grafické znázornenie pomerového ukazovateľa miera pozitívne testovaných osôb v grafe č. 1.

Zverejňovanie údajov o najvyššom počte pozitívne testovaných osôb v určitých časových okamihoch v Bratislavskom kraji predstavuje evidentne zavádzajúcu informáciu z pohľadu rizika šírenia sa koronavírusu. Rovnako to platí aj pre ostatné kraje. Bez poznania ďalších potrebných súvislostí pôsobí takáto informácia na bežného občana výrazne zmätočne. Podľa nej mal byť Bratislavský kraj v daných časových okamihoch najrizikovejším krajom a Košický kraj mal patriť k tým bezpečnejším, pokiaľ išlo o šírenie koronavírusu. Pohľad na graf č. 1 poskytuje čitateľovi inú, ale z nášho pohľadu správnejšiu informáciu oproti predchádzajúcemu tvrdeniu.

Graf č. 1: Miera pozitívne testovaných osôb (v %) vo vybraných dňoch podľa krajov

Zdroj: https://github.com/Institut-Zdravotnych-Analyz/covid19-data/tree/main/PCR_Tests, dátum dostupnosti 30. 12. 2021, vlastné spracovanie

Riziko šírenia koronavírusu v jednotlivých krajoch, znázornené na grafe č. 1 pomocou pomerového ukazovateľa miera pozitívne testovaných osôb, predstavuje realite bližšiu informáciu, ako je tá, ktorú poskytuje absolútny ukazovateľ počtu pozitívne testovaných osôb. Naznačené používanie nepresných a často až zavádzajúcich informácií o vývoji epidemickej situácie má nepriaznivý vplyv na psychológiu správania ľudí v jednotlivých regiónoch.

Menší problém to spôsobuje v kraji, ktorý sa na základe zverejňovania absolútneho ukazovateľa počet pozitívne testovaných osôb považuje za rizikovejší, ako v skutočnosti je. Naopak, veľmi nebezpečná situácia vzniká, keď sa podľa pravidelne zverejňovaného absolútneho ukazovateľa považuje situácia v regióne za menej rizikóvu, ako v skutočnosti je. Vedie to k podceňovaniu reálne horšej epidemickej situácie a znižovaniu obozretnosti ľudí. Samostatné údaje o počte pozitívne testovaných PCR testami podľa krajov bez ďalšej dôležitej informácie o celkovom počte vykonaných PCR testov sú stále každodennou súčasťou prehľadu o aktuálnej situácii s covid-19 na Slovensku. To svedčí o tom, že psychologická rovina naznačeného používania nesprávnych informácií o epidemickej situácii sa pri jej manažovaní stále podceňuje.

V záujme získania korektnej informácie o miere šírenia rizika koronavírusu v krajoch a okresoch je potrebné k pravidelne zverejňovaným údajom o prírastku a celkovom počte novo pozitívne testovaných pridať aj údaj o počte vykonaných PCR testov. Pre bežného človeka by však bolo užitočnejšie, aby sa ako jeden z rozhodujúcich údajov na hodnotenie epidemickej situácie priamo zverejňoval údaj o miere novo pozitívne testovaných podľa krajov, prípadne aj podľa okresov.

3. UŽ SA NEPOUŽÍVA PÔVODNÝ UKAZOVATEĽ REPRODUKČNÉ ČÍSLO

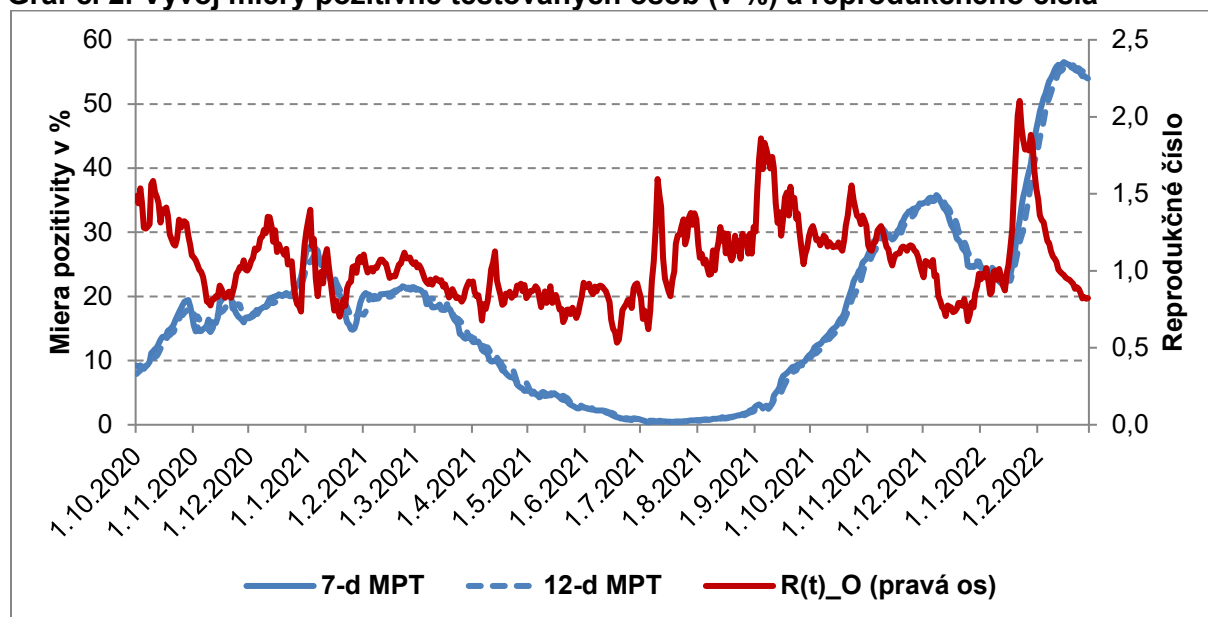
Viac ako rok sa pri hodnotení vývoja epidemickej situácie oficiálne dost' frekventovane používal ukazovateľ reprodukčné číslo šírenia koronavírusu. Zjednodušený výpočet tohto ukazovateľa je: ²

$$R(t) = \frac{\sum_{i=d-7}^{d-1} Y(i)}{\sum_{i=d-6}^{d-12} Y(i)} \quad (1)$$

kde d znamená deň, za ktorý sa reprodukčné číslo vypočítava, $Y(i)$ predstavuje počet pozitívne testovaných osôb za príslušné dni.

Už v polovici roka 2021 sme spomenuli určitú záhadnosť ukazovateľa reprodukčné číslo napríklad v [1]. Postupne sa začala ukazovať ako veľmi diskutabilná samotná vypovedacia schopnosť tohto ukazovateľa a v poslednej štvrtine roka 2021 sa tento ukazovateľ dokonca prestal oficiálne používať. ³ Prvá výhrada k danému ukazovateľu spočíva v tom, že je založený na počte pozitívne testovaných osôb, ktorý do určitej miery skresľuje výslednú informáciu o riziku koronavírusu v dôsledku existujúcej asymetrie vo vzťahu k meniacemu sa počtu vykonávaných PCR testov. Zásadnejšou výhradou je však nelogický vzťah medzi hodnotami oficiálneho ukazovateľa reprodukčné číslo a hodnotami ukazovateľa miera pozitívne testovaných osôb. Za normálnych okolností by sa dala predpokladať logická korelácia medzi uvedenou dvojicou ukazovateľov. Reálne údaje však tento predpoklad nepotvrdili.

Graf č. 2: Vývoj miery pozitívne testovaných osôb (v %) a reprodukčného čísla



Zdroj údajov: <https://korona.gov.sk/koronavirus-na-slovensku-v-cislach/>, **vlastné výpočty, dátum dostupnosti 28. 2. 2022**

² Daná metodika vychádza zo zjednodušenej verzie nemeckého Inštitútu Roberta Kocha a je v súlade napr. s metodikou MZ ČR na výpočet indexu rizika COVID-19 na stránke https://www.mzcr.cz/wp-content/uploads/2021/01/Priloha03_Metodick%C3%BDPopisIndexRizika_verze2.0.pdf. Treba vedieť, že uvedený výpočet ukazovateľa reprodukčné číslo je do určitej miery zjednodušený a je zaťažený rôznymi neurčitostami, ako je dĺžka inkubačného času a infekčnosťou vírusu.

³ Nelogický vývoj hodnôt oficiálneho ukazovateľa reprodukčné číslo si zrejme uvedomili aj zodpovední v Národnom centre zdravotných informácií a v priebehu septembra 2021 ho prestali zverejňovať na stránke <https://korona.gov.sk/koronavirus-na-slovensku-v-cislach/>. Oficiálne zdôvodnenie ani po oslovení zodpovedných inštitúcií nie je známe, len sa možno domnievať, že zodpovedné inštitúcie si už tiež uvedomili zavádzajúcu vypovedaciu schopnosť tohto ukazovateľa.

Z porovnania vývoja hodnôt ukazovateľov sedemdňová a dvanásťdňová miera pozitívne testovaných osôb (7-d MPT a 12-d MPT) k hodnotám ukazovateľa reprodukčné číslo šírenia koronavírusu v grafe č. 2 vo všeobecnosti vyplýva, že riziko šírenia koronavírusu a dynamika jeho šírenia boli v určitých časových úsekoch v dosť výraznom protiklade. Tak napríklad na vrchole druhej vlny epidemickej situácie na začiatku januára 2021 (7-d MPT aj 12- MPT atakovala 30 %) sa hodnota oficiálneho reprodukčného čísla pohybovala aj pod hodnotou 1, kým v priebehu júla a augusta 2021, keď bola miera pozitívne testovaných pod hodnotou 0,5 %, sa hodnota reprodukčného čísla pohybovala takmer výlučne nad hodnotou 1. Problematické je obhájiť protichodný vývoj hodnôt porovnávaných ukazovateľov aj na prelome mesiacov január a február v roku 2022. Dosť ťažko sa hľadá logické vysvetlenie na uvedený protirečivý vzťah vzájomne sa podmieňujúcich ukazovateľov, ktoré vyjadrujú riziko šírenia koronavírusu a dynamiku šírenia koronavírusu.

Pri rastúcej miere pozitívne testovaných osôb sa vo všeobecnosti očakáva, že je spojená s rýchlejšim šírením koronavírusu, a naopak. Preto je potrebné uvažovať, akým spôsobom by bolo možné vytvoriť ukazovateľ reprodukčné číslo, aby jeho hodnoty korešpondovali so šírením koronavírusu, zisťovaným pomocou ukazovateľa miera pozitívne testovaných osôb.

Riešením teda nie je úplné zrušenie ukazovateľa reprodukčné číslo bez akejkoľvek náhrady podobným, realitu viac odrážajúcim ukazovateľom. Možností môže byť viac a jednu z nich predstavíme v ďalšej časti príspevku.

Jednou z možností, ako vytvoriť aj reálne použiteľné reprodukčné číslo šírenia koronavírusu, je založiť jeho konštrukciu na výpočte a vážení pomocou miery pozitívne testovaných osôb v percentách.

Výpočet navrhovaného ukazovateľa vážené reprodukčné číslo je:

$$R(v) = \frac{\sum_{i=d}^{d-7} m(i)}{\sum_{i=d-6}^{d-12} m(i)} * v(i) \quad (2)$$

kde d znamená deň, za ktorý sa reprodukčné číslo vypočítava, $m(i)$ predstavuje mieru pozitívne testovaných osôb v percentách za príslušné dni a $v(i)$ prepočítací váhový koeficient na príslušný deň.

Nový ukazovateľ vážené reprodukčné číslo šírenia koronavírusu sa vytvára z údajov za posledných 12 dní, preto ako východisko pre prepočítací váhový koeficient použijeme ukazovateľ miera pozitívne testovaných za posledných 12 dní.

Výpočet ukazovateľa vážené reprodukčné číslo obsahuje nasledujúce kroky:

1. Východiskom na výpočet váženého reprodukčného čísla pre príslušný deň podľa vzťahu (2) sú hodnoty priemernej miery pozitívne testovaných osôb počas posledných dvanásť dní.⁴
2. Takto vypočítané reprodukčné číslo sa následne preváži príslušným prepočítacím váhovým koeficientom (v).

⁴ *Dá sa overiť, že nami zvolený prístup poskytuje relatívne vyrovnanejšie hodnoty časového radu váženého reprodukčného čísla šírenia koronavírusu ako pri výpočte na základe počtu pozitívne testovaných osôb.*

3. Prepočítací váhový koeficient (v) sa určí na základe dosiahnutej priemernej hodnoty miera pozitívne testovaných za posledných dvanásť dní a to pomocou lineárnej interpolácie.

Návrh hodnôt prepočítacieho váhového koeficienta vychádza z nasledujúcej úvahy:

- V prípade hodnôt dvanásťdňovej priemernej miery pozitívne testovaných osôb od 0,1 až do hodnoty jedna, bude týmto hodnotám zodpovedať aj prepočítací váhový koeficient na príslušný deň.
- Pri hodnotách ukazovateľa dvanásťdňová miera pozitívne testovaných osôb nad jedna sa pri jeho zvýšení o jedno percento zvýši prepočítací váhový koeficient lineárne o jednu desatinu.

Pre lepšiu názornosť je spôsob stanovovania prepočítacieho váhového koeficienta demonštrovaný v tabuľke č. 2. Hodnoty pre prepočítací koeficient, stanovené detailnejšie (po jednej desatinke) považujeme za vhodnejší postup, ako ich stanovenie pomocou menej detailnej škály. Na základe detailnejšej škály získame hladší rad hodnôt váženého reprodukčného čísla.

Tabuľka č. 2: Navrhnutý postup stanovovania prepočítacieho váhového koeficienta

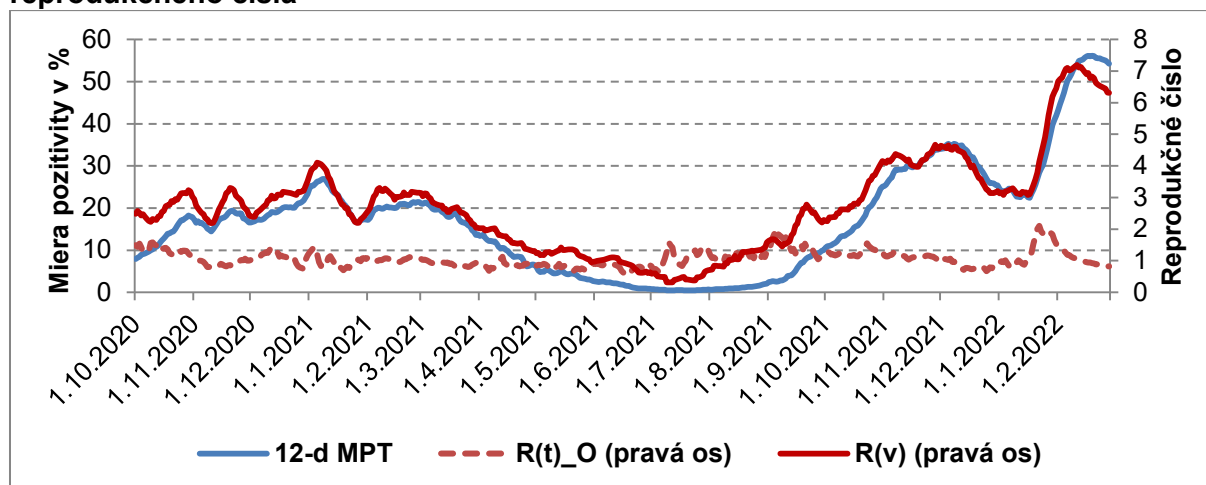
12-d MPT	0,1-1,0	1,1-9,9	10,0-10,9	11,0-19,9	20,0-20,9	21,0-29,9	30,0-30,9	31,0-39,9
$v(i)$	0,1, ... 1,0	1,1, ... 1,9	2,0	2,1, ... 2,9	3,0	3,1, ... 3,9	4,0	4,1, ... 4,9

Zdroj: vlastné spracovanie

Poznámka: Pre hodnoty dvanásťdňovej miery pozitívne testovaných (12-d MPT) od 40 percent sa váhový koeficient (v_i) stanoví pomocou lineárnej interpolácie rovnako ako pre predchádzajúce kategórie.

Na základe vypočítaného váženého reprodukčného čísla šírenia koronavírusu podľa opísaného postupu získame ukazovateľ, ktorý má podľa nás lepšiu vypovedaciu schopnosť ako pôvodne používaný oficiálny ukazovateľ reprodukčné číslo šírenia koronavírusu. Inými slovami, navrhovaný nový ukazovateľ vážené reprodukčné číslo by mal logicky reagovať tak, že pri vyššom riziku šírenia koronavírusu (t. j. pri vyššej miere pozitivity) by mala byť aj rýchlosť šírenia koronavírusu (vyjadrená pomocou ukazovateľa vážené reprodukčné číslo) vyššia, a naopak. Rozdiel medzi pôvodným reprodukčným číslom a váženým reprodukčným číslom je dosť zrejmy z grafu č. 3.

Graf č. 3: Vývoj miery pozitívne testovaných osôb (v %), pôvodného a váženého reprodukčného čísla



Zdroj údajov: <https://korona.gov.sk/koronavirus-na-slovensku-v-cislach/>, **vlastné výpočty, dátum dostupnosti 28. 2. 2022**

Hodnoty nového váženého reprodukčného čísla v celej analyzovanej histórii dosť jednoznačne korelujú s hodnotami miery pozitívne testovaných osôb. Sú aj výrazne variabilnejšie ako hodnoty pôvodného oficiálneho reprodukčného čísla. Nové vážené reprodukčné číslo dosiahlo v hodnotenom období maximálnu hodnotu 7,2 na konci prvej dekády februára 2022, keď vrcholila štvrtá vlna pandémie koronavírusu. Maximom tohto ukazovateľa v druhej vlne pandémie v prvom januárovom týždni v roku 2021 bola hodnota 4,1 a v tretej vlne na konci novembra 2021 hodnota 4,7. Najnižšiu hodnotu 0,3 dosiahlo nové vážené reprodukčné číslo v dňoch 9. 7. 2021 až 12. 7. 2021.

Ukazovateľ pôvodného oficiálneho reprodukčného čísla by dosiahol v čase vrcholu tretej vlny epidemickej situácie na konci prvého decembrového týždňa 2021 (pri sedemdnovej miere pozitívne testovaných osôb okolo 35 %) hodnotu skôr pod jedna a na vrchole štvrtej vlny pandémie okolo 10. februára 2022 (pri sedemdnovej miere pozitívne testovaných osôb okolo 55 %) hodnotu jedna. Inými slovami, jeden človek infikovaný koronavírusom by pri enormnom zvýšení rizika koronavírusu nakazil len jednu osobu, čo nereflektuje realitu. Podľa nového navrhovaného ukazovateľa by na vrchole tretej vlny pandémie koronavírusu nakazil jeden človek infikovaný koronavírusom až takmer 5 osôb a na vrchole štvrtej vlny až viac ako 7 osôb. Podľa nášho názoru rýchlosť šírenia koronavírusu identifikovaná podľa váženého reprodukčného čísla celkom korešponduje s rizikom šírenia sa koronavírusu v hodnotených kritických obdobiach.

Navrhovaný ukazovateľ nové vážené reprodukčné číslo na rozdiel od pôvodného oficiálneho ukazovateľa reprodukčné číslo poskytuje celkom užitočné a aj reálnejšie sa javiace informácie o rýchlosti šírenia sa koronavírusu. Určite v čase vrcholiaceho šírenia sa koronavírusu vierohodnejšie pôsobí informácia, že jeden pozitívne testovaný človek môže nakaziť až takmer piatich, resp. viac ako sedem ľudí, ako keby sa tvrdilo, že jeden pozitívne testovaný človek môže nakaziť v takej kritickej epidemickej situácii len jedného človeka.

Navrhovaný ukazovateľ vážené reprodukčné číslo predstavuje určitý pokus vytvoriť funkčný ukazovateľ na spoľahlivé sledovanie rýchlosti šírenia koronavírusu. Náš návrh považujeme za jeden z možných prístupov k riešeniu daného problému a je to zároveň príspevok do diskusie o hľadaní účinných metód a nástrojov na reálne hodnotenie epidemickej situácie.

4. STRUČNÉ ZHODNOTENIE OSTATNÝCH VLŇ PANDÉMIE KORONAVÍRUSU PODĽA VYBRANÝCH UKAZOVATEĽOV

Už takmer dvojročná skúsenosť s hodnotením epidemickej situácie potvrdila, že pre širokú verejnosť je jej hodnotenie viac zrozumiteľné, ak je realizované pomocou menšieho počtu vhodných ukazovateľov. My budeme stručne hodnotiť vývoj epidemickej situácie pomocou niekoľkých relevantných ukazovateľov. Za také považujeme ukazovateľ miera pozitívne testovaných osôb, ukazovateľ počet hospitalizovaných pacientov v nemocniciach s ochorením covid-19 a ukazovateľ počet úmrtí na covid-19.

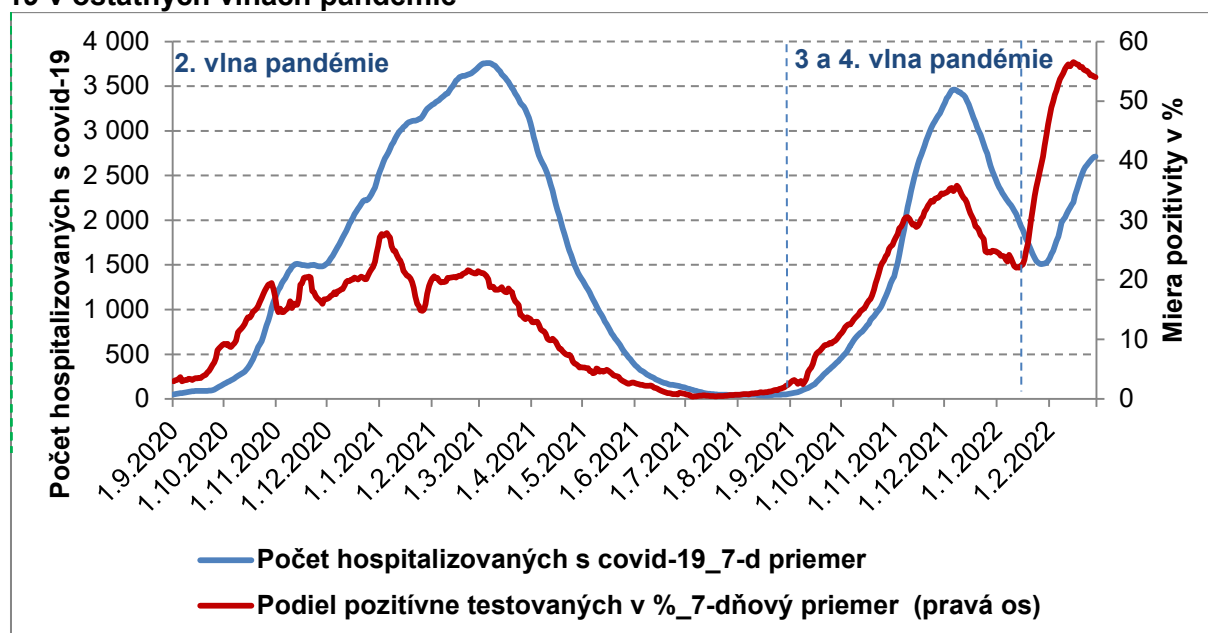
Naše hodnotenia vychádzajú z oficiálnych údajov, bežne dostupných na portáloch <https://korona.gov.sk/koronavirus-na-slovensku-v-cislach/> a <https://covid-19.nczisk.sk/sk> za obdobie od začiatku septembra 2020 do konca februára 2022. Sme si vedomí určitého zjednodušenia pri našich hodnoteniach, pretože napríklad pri

čísloch o úmrtiach na covid-19 sa nebudeme podrobnejšie zaoberať prípadnými diagnózami zverejnených úmrtí na covid-19, ani ich vekovou skladbou a podobne.

Uvedené ukazovatele pokrývajú podľa nás tie najdôležitejšie oblasti, ktoré by nemali chýbať pri sledovaní a hodnotení vývoja epidemickej situácie. Veľkosť rizika šírenia sa koronavírusu veľmi úzko súvisí s potenciálne zvýšenou chorobnosťou a následne aj s napĺňaním nemocničných lôžok. Celý proces šírenia koronavírusu v mnohých prípadoch dospieva až do smutného konca, k úmrtiu. Obmedzíme sa len na uvedené ukazovatele, ale zvažovať možno aj použitie funkčného ukazovateľa na spoľahlivé sledovanie rýchlosti šírenia koronavírusu v podobe novo navrhovaného ukazovateľa vážené reprodukčné číslo a niekto by možno privítal aj informáciu o miere vakcinácie populácie proti covidu-19.

Na porovnávanie vývoja rizika šírenia sa koronavírusu použijeme sedemdňové priemery miery pozitívne testovaných osôb z dôvodu získania vyrovnaných časových radov ich hodnôt. Aj pri ukazovateli počet hospitalizovaných s covidom-19 použijeme ich sedemdňový priemer. V záujme plastickejšieho grafického priblíženia vývoja hodnôt vybraných ukazovateľov počas druhej a tretej vlny pandémie koronavírusu upriamime pozornosť na obdobie od začiatku septembra v posledných dvoch rokoch. Štvrtá vlna pandémie, ktorej vrchol bol zaznamenaný v polovici februára 2022, je z pohľadu vybraných ukazovateľov špecifická aj v dôsledku výskytu novej mutácie koronavírusu omikron.

Graf č. 4: Vývoj miery pozitívne testovaných (v %) a počtu hospitalizovaných s covidom-19 v ostatných vlnách pandémie



Zdroj údajov: <https://korona.gov.sk/koronavirus-na-slovensku-v-cislach/>, vlastné výpočty, dátum dostupnosti 28. 2. 2022

Mesiac september možno celkom oprávnené považovať za začiatok druhej vlny pandémie v roku 2020, ale aj tretej vlny pandémie v roku 2021. Z grafu č. 4 je zrejme, že v oboch predchádzajúcich rokoch zhruba na konci prvého septembrového týždňa sa začali hodnoty ukazovateľa miera pozitívne testovaných osôb dynamicky zvyšovať a na konci mesiaca sa už pohybovali okolo 10 %. Znamená to, že na konci mesiaca bol pozitívne testovaný na koronavírus jeden človek spomedzi desiatich testovaných

pomocou PCR testov. Prekročenie tejto hranice a predovšetkým permanentne stúpajúci trend rizika šírenia sa koronavírusu boli jednoznačným signálom, že je potrebné uvažovať o radikálnejších vládných opatreniach proti ďalšiemu zhoršovaniu epidemickej situácie.

V druhej vlne z dôvodu výrazne sa zhoršujúcej epidemickej situácie od 1. októbra 2020, na základe odporúčania ústredného krízového štábu vláda vyhlásila núdzový stav. Ten bol sprevádzaný celým radom opatrení hlavne smerom k zníženiu mobility obyvateľstva. V druhej polovici októbra 2020 vláda schválila návrh na prípravu a vykonanie celoštátneho plošného testovania obyvateľstva na prítomnosť ochorenia covid-19.

Protiepidemické opatrenia zo začiatku októbra 2020 prispeli k zastaveniu rastu miery pozitívne testovaných na niekoľko dní na prelome mesiacov októbra a novembra. Celoplošné testovanie nemalo očakávaný efekt, lebo hneď od prvých dní novembra až do konca druhej dekády novembra sa epidemická situácia opäť postupne zhoršovala. Potom po vyše týždňovom zmiernení rizika šírenia koronavírusu sa však od začiatku decembra začala epidemická situácia opäť zhoršovať a vianočné dni odštartovali ešte výraznejšie zhoršovanie, ktoré vyvrcholilo v prvom januárovom týždni v roku 2021. Vtedy bol zaznamenaný vrchol šírenia koronavírusu v rámci druhej vlny pandémie. Najvyššiu dennú mieru pozitívne testovaných osôb pomocou PCR testov sme zaznamenali 1. 1. 2021 (35 %) a 5. 1. 2021 bol zaznamenaný dovtedy najvyšší sedemdňový priemer tohto ukazovateľa (27,8 %).

Tretia vlna pandémie sa javí na Slovensku z pohľadu vývoja miery pozitívne testovaných osôb agresívnejšia ako druhá vlna. Na pomerne výrazné zvyšovanie sa tejto miery od konca prvého septembrového týždňa v roku 2021 zodpovedné inštitúcie prakticky nereagovali, a preto už na začiatku novembra 2021 bolo prekonané maximum sedemdňového priemeru miery pozitívne testovaných osôb z druhej vlny pandémie (keď 4. 11. 2021 predstavovala sedemdňová miera rizika šírenia koronavírusu 28,6 %). Znamená to, že k prekonaniu maxima z druhej vlny došlo s dvojmesačným predstihom.

Po krátkom spomalení rastu hodnôt sedemdňovej miery pozitívne testovaných osôb v polovici novembra 2021 sa začala epidemická situácia na Slovensku opäť zhoršovať a vrchol tretej vlny pandémie sa dosiahol 8. decembra 2021 na úrovni 35,8 %. Zhodou okolností bola v ten istý deň zaznamenaná aj dovtedy najvyššia denná miera pozitívne testovaných osôb pomocou PCR testov (38,3 %). Znamená to, že v tretej vlne pandémie dosiahla vrchol zhruba o mesiac skôr ako v druhej vlne. Odvtedy sa hodnoty sedemdňového priemeru miery pozitívne testovaných osôb postupne znižovali.

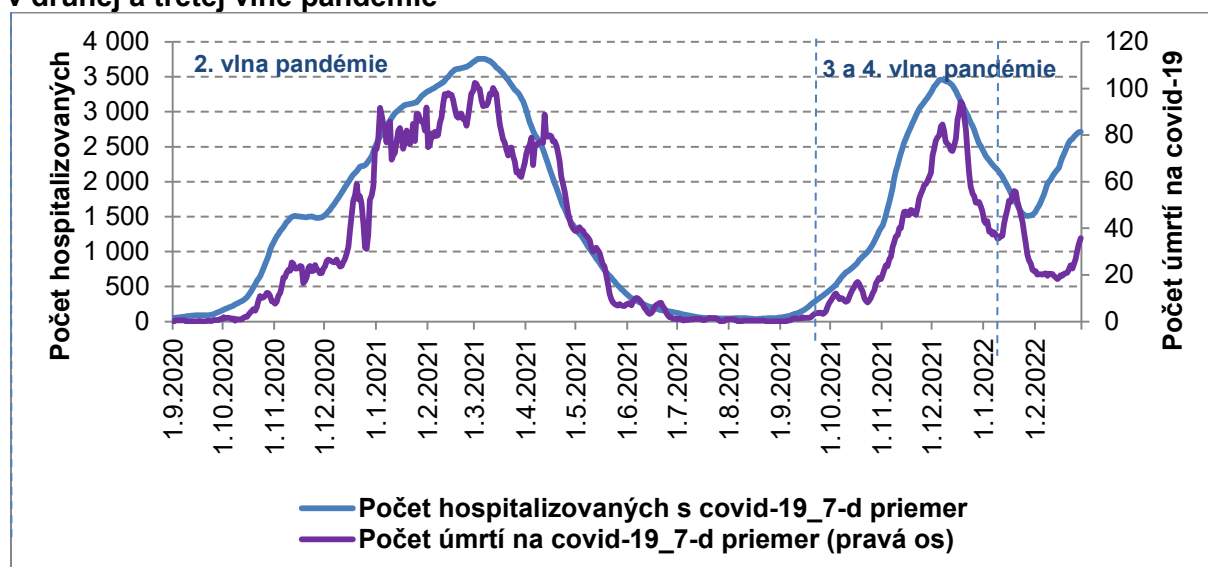
Relatívne priaznivý vývoj epidemickej situácie na Slovensku v prvej polovici januára 2022 mohol byť výsledkom pôsobenia viacerých faktorov. Spomína sa aj dosť výrazná miera premorenia slovenskej populácie, pretože Slovensko patrilo (spolu so Slovinskom, Chorvátskom, Ukrajinou a Bosnou a Hercegovinou) v Európe v podstate od začiatku novembra 2021 medzi malý počet krajín, v ktorých sa medzitýždňová miera pozitívne testovaných osôb pohybovala dlhodobo nad 20 %. Ďalším faktorom mohol byť vyhlásený lockdown pred Vianocami a s ním spojené pomerne prísne protiepidemické opatrenia. Pre mnohých odborníkov bol tento relatívne priaznivý vývoj pandémie koronavírusu na začiatku roka 2022 dosť prekvapujúci vzhľadom na už výrazný výskyt nového variantu covid-19 s označením omikron aj v okolitých krajinách.

Odborníci očakávali, že omikron sa na prelome januára a februára 2022 v plnej sile prejaví aj na Slovensku. Mierne zvyšovanie miery pozitívne testovaných osôb nastalo zhruba už v polovici januára 2022. V priebehu ďalších dní sa začala sedemdňová miera pozitívne testovaných osôb veľmi rýchlo zvyšovať a v priebehu mesiaca sa na vrchole štvrtej vlny pandémie v polovici februára 2022 viac ako zdvojnásobila (zo zhruba 23 % v polovici januára 2022 na vyše 56 % v polovici februára 2022). Naplnili sa predpoklady, že pri omikrone sa prudko zvýši miera pozitívne testovaných osôb v relatívne krátkom čase a po dosiahnutí vrcholu by malo nasledovať tiež výraznejšie znižovanie hodnôt daného ukazovateľa. Na konci februára 2022 bola na Slovensku sedemdňová miera pozitívne testovaných osôb ešte stále veľmi vysoká (54 %), čím sa naša krajina zaraďovala medzi päťicu európskych krajín (spolu so Slovinskom, Estónskom, Faerskými ostrovmi a Islandom) s rizikom šírenia sa koronavírusu nad 50 percent.

Z pohľadu na graf č. 4 vyplýva celkom logická výrazná súvislosť medzi časovými radmi hodnôt miery pozitívne testovaných osôb pomocou PCR testov a hodnôt počtu hospitalizovaných pacientov s covidom-19. Naše porovnanie je založené na sedemdňových priemeroch hodnôt daných ukazovateľov. Najvyššia bola korelácia medzi hodnotami porovnávaných ukazovateľov v tretej vlne pandémie (od 1. 9. 2021 do 15. 1. 2022), keď koeficient korelácie predstavoval $r = 0,95$. Mierne nižšia bola korelácia medzi porovnávanými ukazovateľmi v druhej vlne pandémie (od 1. 9. 2020 do 15. 1. 2021), keď $r = 0,91$, a najnižšia vo štvrtej vlne pandémie (od 15. 1. 2022 do 15. 2. 2022) s $r = 0,4$.

Z pohľadu hospitalizovaných s covidom-19 v priebehu jednotlivých vln pandémie koronavírusu je zaujímavé to, že na ich vrchole boli sedemdňové priemerné počty hospitalizovaných nepriamo úmerné riziku šírenia koronavírusu. Rastúcim hodnotám sedemdňovej miery pozitívne testovaných osôb na vrchole ostatných vln pandémie koronavírusu zodpovedali klesajúce počty hospitalizovaných s covidom-19. Jedným z dôvodov daného zistenia je pravdepodobne aj postupne sa zvyšujúci stupeň premorenia obyvateľov a ich rastúca odolnosť voči nákaze koronavírusom.

Ak sme v predchádzajúcej časti príspevku konštatovali výraznú súvislosť medzi časovými radmi hodnôt miery pozitívne testovaných osôb pomocou PCR testov a hodnôt počtu hospitalizovaných pacientov s covidom-19, tak časové rady hodnôt počtu hospitalizovaných pacientov s covidom-19 a hodnôt počtu úmrtí na covid-19 majú ešte významnejšiu koreláciu. Naše porovnanie je opäť založené na sedemdňových priemeroch hodnôt porovnávaných ukazovateľov od začiatku septembra 2020 do konca februára 2022.

Graf č. 5: Vývoj počtu hospitalizovaných s covidom-19 a počtu úmrtí na covid-19 v druhej a tretej vlne pandémie

Zdroj údajov: <https://korona.gov.sk/koronavirus-na-slovensku-v-cislach/>, vlastné výpočty, dátum dostupnosti 28. 2. 2022

Na celej hodnotenej trajektórii predstavoval koeficient korelácie medzi uvedenými ukazovateľmi hodnotu $r = 0,93$. Po vrchol počtu hospitalizovaných s covidom-19 v druhej vlne pandémie (od 1. 9. 2020 do 5. 3. 2021) predstavoval koeficient korelácie medzi uvedenými ukazovateľmi hodnotu $r = 0,91$, v tretej vlne pandémie (od 1. 9. 2021 do 6. 12. 2021) hodnotu $r = 0,98$. Vo štvrtej vlne zatiaľ registrujeme medzi porovnávanými ukazovateľmi zápornú koreláciu, hlavne v dôsledku relatívne nízkej smrtnosti omikronu.

Vývoj počtu hospitalizovaných pacientov s covidom-19 a hodnôt počtu úmrtí na covid-19 bol v druhej a tretej vlne pandémie relatívne symetrický. Vyššiemu počtu hospitalizovaných s covidom-19 na vrchole druhej vlny pandémie zodpovedal aj vyšší počet úmrtí na covid-19 a nižšiemu počtu hospitalizovaných s covidom-19 na vrchole tretej vlny pandémie zodpovedal aj nižší počet úmrtí na covid-19 (v priemere 3632 hospitalizovaných s covidom-19 zo 14. 3. 2021 a v priemere 102 úmrtí z 1. 3. 2021 oproti v priemere 3435 hospitalizovaných s covidom-19 z 9. 12. 2021 a v priemere 94 úmrtí na covid-19 z 18. 12. 2021). Na konci februára boli hodnoty dvoch porovnávaných ukazovateľov už zhruba na vrchole a boli nižšie (v priemere 2 712 hospitalizovaných s covidom-19 a v priemere 87 úmrtí na covid-19) ako v predchádzajúcich vlnách pandémie.

Podľa dostupných údajov aj za prvé dni v marci 2022 sa ukázalo, že koncom februára bol dosiahnutý vrchol sedemdňového priemeru počtu hospitalizácií s covidom-19 v ostatnej vlne pandémie. Po pár dňoch sa začal znižovať aj sedemdňový priemer počtu úmrtí na covid-19. Pre štvrtú vlnu pandémie je charakteristická zhruba trojtýždňová oneskorená reakcia rastu sedemdňového priemeru počtu úmrtí na covid-19 (od 19. 2. 2022) oproti začiatku rastu sedemdňového priemeru počtu hospitalizácií s covidom-19 (od 28. 1. 2022).

Zdá sa, že aj pomocou obmedzeného počtu vhodne vybraných a vzájomne konzistentných ukazovateľov možno poskytnúť celkom užitočné informácie o globálnom vývoji epidemickej situácie. Je pochopiteľné, že na rôznych stupňoch

manažovania boja proti šíreniu koronavírusu je vhodné použiť aj ďalšie doplňujúce ukazovatele.

5. ZÁVER

Pravidelne zverejňované údaje o prírastku a celkovom počte novo pozitívne testovaných osôb podľa krajov na webovej stránke <https://korona.gov.sk/koronavirus-na-slovensku-v-cislach/> poskytuje nepresné až zavádzajúce informácie. Môže to mať dosť negatívny na správanie obyvateľov, hlavne keď daný ukazovateľ vyhodnocuje epidemickú situáciu vo vzťahu k iným regiónom ako menej rizikový, ako je v skutočnosti. Možno to vyhodnotiť aj tak, že sa stále podceňuje psychologická rovina používania nesprávnych informácií o vývoji epidemickej situácie pri jej manažovaní.

Ukazovateľ reprodukčné číslo patril k jedným z najfrekvencovanejších ukazovateľov pri hodnotení epidemickej situácie, ale od septembra 2021 sa prestal bez akéhokoľvek zdôvodnenia oficiálne zverejňovať. Navrhujeme vytvoriť reálne použiteľné reprodukčné číslo šírenia koronavírusu, ktoré by bolo založené na vážení pomocou miery pozitívne testovaných.

Trojica ukazovateľov miera novo pozitívne testovaných osôb, počet hospitalizovaných s covidom-19 a počet úmrtí na covidom-19 dokáže poskytnúť celkom užitočné informácie aj celkom plastický obraz o globálnom vývoji epidemickej situácie. Na základe uvedenej trojice je možné zachytiť aj dôležité rozdiely medzi jednotlivými vlnami cyklicky sa vyvíjajúcej pandémie koronavírusu.

Efektívne hodnotenie a manažovanie epidemickej situácie predpokladá, aby sa na tento účel používal systém vhodných a hlavne navzájom konzistentných ukazovateľov. Je potrebné dbať na to aby jednotlivé ukazovatele, používané pri sledovaní a hodnotení epidemickej situácie, nevyvolávali pochybnosti o tom, ako reflektujú realitu.

LITERATÚRA

- [1] CÁR, M.: Záhadné reprodukčné číslo šírenia koronavírusu, In: Slovenská štatistika a demografia, 2021, č.3, s. 94 – 99.
- [2] CÁR, M.: Otázky týkajúce sa hodnotenia šírenia koronavírusu. In: Forum Statisticum Slovaca, 2020, č. 2, s. 1- 10.
- [3] CÁR, M.: Dôležité empirické poznatky z hodnotenia epidemickej situácie na Slovensku. In: Slovenská štatistika a demografia, 2021, č.4, s. 20 – 37.
- [4] Parag, K. V. and collective: Are epidemic growth rates more informative than reproduction numbers? doi: <https://doi.org/10.1101/2021.04.15.21255565>.
- [5] Stankovičová, I. – Boďa, M.: Sledovať kľzavý medián počtu pozitívne testovaných osôb na koronavírus je nekorektné. In.: Slovenská štatistika a demografia, 2020, č. 4, s. 46 – 47.

RESUMÉ

Príspevok poukazuje na každodenné používanie nepresného absolútneho ukazovateľa počet novo pozitívne testovaných osôb, ktorý môže byť až zavádzajúci a nebezpečný pre správanie obyvateľstva v boji proti šíreniu koronavírusu. Zamýšľa sa nad sfunkčnením oficiálne vyradeného, ale predtým veľmi frekvencovaného ukazovateľa reprodukčné číslo šírenia koronavírusu. Prináša pohľad na ostatné vlny pandémie a ich porovnanie prostredníctvom obmedzeného počtu relevantných ukazovateľov. Predpokladané ďalšie vlny epidémie koronavírusu sú výzvou na

dôsledné odborné zhodnotenie doterajších postupov a vytvorenie dostatočne účinného systému vhodných a hlavne navzájom konzistentných ukazovateľov na korektné hodnotenie epidemickej situácie aj v budúcnosti.

RESUME

The paper points to the daily use of an inaccurate absolute indicator of the number of newly tested persons, which can be misleading and dangerous for the behaviour of the population in the fight against the spread of coronavirus. It has questions regarding the functioning of the officially discarded, but previously very frequent indicator, the reproductive number of the spread of coronavirus. It provides an overview of other pandemic waves and compare them through a limited number of relevant indicators. The expected further waves of the coronavirus epidemic are a challenge for a thorough professional evaluation of the current procedures and the creation of a sufficiently effective system of suitable and especially mutually consistent indicators for a correct assessment of the future epidemiological situation.

PROFESIJNÝ ŽIVOTOPIS

Ing. Mikuláš Cár, PhD., je absolventom Ekonomickej univerzity v Bratislave. Dlhodobu sa venuje analýzam sociálno-ekonomických javov a procesov. Má pomerne bohatú publikačnú činnosť.

KONTAKT

mikulas.car@gmail.com

Milan TEREK
Vysoká škola manažmentu

ANALÝZA KAUZÁLNYCH VZŤAHOV MEDZI PREMENNÝMI

ANALYSIS OF CAUSAL RELATIONS BETWEEN VARIABLES

ABSTRAKT

Overovanie a odhadovanie predpokladaných kauzálnych vzťahov medzi premennými sa považuje za centrálny problém. Cieľom článku je poskytnúť prehľad o existujúcich postupoch a metódach, ktoré možno v tejto oblasti uplatniť. Sú v ňom opísané ich základné črty. Možnosti ich využívania sú ilustrované na jednoduchých príkladoch. To môže uľahčiť proces výberu vhodných metód pri riešení konkrétnych problémov overovania a odhadovania predpokladaných kauzálnych vzťahov medzi premennými.

ABSTRACT

Verifying and estimating the presumed causal relations between the variables is considered a central problem. The aim of the article is to provide an overview of the existing procedures and methods applicable in this area. The article describes their basic features and illustrates the possibilities of their use by simple examples. This can facilitate the process of selecting appropriate methods for solving specific problems of verification and the estimation of the presumed causal relationships between variables.

KLÚČOVÉ SLOVÁ

kauzálne vzťahy medzi premennými, Simpsonov paradox, analýza ciest, štruktúrne rovnicové modely, kauzálna indukcia.

KEY WORDS

Causal relations between variables, Simpson paradox, path analysis, structural equation models, causal inference.

1. ÚVOD

Príspevok je venovaný možnostiam overovania a odhadovania predpokladaných kauzálnych vzťahov medzi premennými. Analýzu kauzálnych vzťahov medzi premennými možno považovať za centrálny problém. Opíšeme základné črty postupov a metód, ktoré možno v tejto oblasti uplatniť. Možnosti ich aplikácie budeme ilustrovať na jednoduchých príkladoch.

Všimneme si najprv pojmy asociácia a kauzalita. Keď určité hodnoty jednej premennej majú tendenciu objavovať sa spolu s určitými hodnotami druhej premennej, hovoríme že medzi premennými je asociácia alebo spojenie (*association*). Definícia kauzality, ktorú by všeobecne akceptovali štatistickí, filozofi aj vedci z iných oblastí, nie je známa. Uvedieme najprv jednu, trochu metaforickú definíciu, ktorá ale poskytuje veľmi dobrú intuitívnu predstavu o kauzalite [10, s. 5]: „Premenná X má kauzálny vplyv na premennú Y , keď Y počúva X a rozhoduje o svojej hodnote na základe toho, čo počuje“. Uvedieme aj presnejšiu definíciu. Zápisom $X \rightarrow Y$ označíme, že premenná X má kauzálny vplyv na premennú Y . Kauzalitu možno potom definovať takto. Keď $X \rightarrow Y$, potom keď sa zmení X , zmení sa rozdelenie Y [3]. Kauzálne vzťahy sú najčastejšie asymetrické. Jedna premenná ovplyvňuje druhú, ale nie opačne.

Samotná asociácia neimplikuje kauzalitu. Kauzálny vzťah medzi premennými implikuje asociáciu, ale opačná implikácia neplatí. Keď je medzi premennými asociácia, môže ale nemusí medzi nimi byť kauzálny vzťah. V [7] je na s. 44 – 45, uvedený zaujímavý príklad. Uvažuje sa o dvoch premenných – počet mentálne postihnutých osôb pripadajúcich na 10 000 obyvateľov v Spojenom kráľovstve a počet vydaných licencií na rádiový prijímač v Spojenom kráľovstve. Sú známe hodnoty týchto premenných v rokoch 1924 až 1937. Skúmal sa lineárny vzťah medzi závisle premennou počet mentálne postihnutých osôb pripadajúcich na 10 000 obyvateľov v Spojenom kráľovstve a nezávisle premennou počet vydaných licencií na rádiový prijímač v Spojenom kráľovstve. Ukázalo sa, že vzťah je štatisticky významný. Koeficient determinácie vyšiel 0,9842. To znamená, že 98,42 % variability závisle premennej možno vysvetliť lineárnym vzťahom medzi počtom mentálne postihnutých osôb pripadajúcich na 10 000 obyvateľov v Spojenom kráľovstve a počtom vydaných licencií na rádiový prijímač v Spojenom kráľovstve. Je zrejmé, že nejde o kauzálny vzťah. Možno jednoducho povedať, že dôvodom štatistického vzťahu medzi týmito premennými je to, že sú spojené monotónnosťou (*monotonically related*).¹

Všeobecne, vzťah medzi premennými možno považovať za kauzálny, keď sú splnené aspoň tieto tri podmienky [4, s. 288]:

- medzi premennými je asociácia,
- príčina predchádza v čase následok,
- boli vylúčené alternatívne vysvetlenia asociácie.

Prvým predpokladom kauzálneho vzťahu je asociácia medzi premennými. Napríklad vzťah medzi rakovinou pľúc a fajčením v regresných analýzach je veľmi silný. To je však len jeden z predpokladov. Je prirodzené, že predpokladaná príčina musí v kauzálnom vzťahu predchádzať následok. Napríklad ľudia často ochorejú na rakovinu pľúc po rokoch fajčenia. Keď je výskumná štúdia experimentálna², možno časovú následnosť v experimente fixovať. V štúdiách založených na pozorovaniach to nie je možné. Vylúčenie alternatívnych vysvetlení asociácie je obyčajne najzložitejší problém. V štúdiách založených na pozorovaniach zvyčajne nie je zložitá nájsť asociácie medzi premennými. Tieto asociácie však možno často vysvetliť inými premennými, ktoré v štúdií neboli merané. Napríklad v mnohých medicínskych štúdiách našli asociáciu medzi pitím kávy a pravdepodobnosťou infarktu myokardu. Keď však v štúdiách zohľadnili aj iné premenné, ako napríklad zamestnanie, úroveň stresu a pod., asociácia medzi pitím kávy a pravdepodobnosťou infarktu sa stratila alebo výrazne zmenšila. Preto je splnenie tejto podmienky kauzálneho vzťahu najviac problematické. Môžeme sa nazdávať, že sme vylúčili všetky iné vysvetlenia asociácie, ale istí si nemôžeme byť nikdy. Preto nemôžeme nikdy dokázať, že jedna premenná je v kauzálnom vzťahu s druhou. Hypotézu o kauzalite môžeme len vyvrátiť tým, že ukážeme, že empirický dôkaz je v rozpore aspoň s jedným z troch predpokladov kauzality [3, s. 303].

Okrem troch uvedených podmienok existencie kauzality medzi premennými niektorí autori uvádzajú aj niektoré ďalšie podmienky ktoré podporujú indície o existencii kauzálneho vzťahu medzi premennými. V [9] sa na s. 301-302 uvádza aj podmienka

¹ Dve postupnosti čísiel sú spojené monotónnosťou, keď s rastom jednej postupnosti druhá postupnosť rastie alebo klesá ([7, s. 45]).

² O experimentálnej štúdií a štúdií založenej na pozorovaniach pozri v [16, s. 19 - 20].

konzistencie vzťahu medzi premennými. Napríklad veľké množstvo štúdií v rozličných krajinách prezentuje vzťah medzi rakovinou pľúc a fajčením. Charakter vzťahu je v rozličných štúdiách rovnaký. Napríklad vo všetkých štúdiách vychádza, že ľudia, ktorí fajčia viac cigariet denne alebo fajčia dlhší čas, majú rakovinu pľúc častejšie. Ďalšou podmienkou je, aby predpokladaná príčina bola možná. Napríklad experimenty so zvieratami ukazujú, že decht z cigariet spôsobuje rakovinu. Pri splnení všetkých podmienok sú indície o existencii a charaktere kauzálneho vzťahu často veľmi silné, ale obyčajne nie také silné ako na základe výsledkov realizácie dobre navrhnutého experimentu.³

Najprv si všimneme možnosti overovania a odhadovania predpokladaných kauzálnych vzťahov medzi jednou závisle premennou a jednou alebo viacerými nezávisle premennými. Tu sa zameriame na základný problém – skúmanie alternatívnych vysvetlení asociácie - a na jeden špecifický problém – Simpsonov paradox. Potom budeme študovať predpokladané kauzálne vzťahy v modeli systému kauzálnych vzťahov, v ktorom môžu byť premenné, o ktorých sa predpokladá, že sú následkom iných premenných a zároveň ovplyvňujú nejaké ďalšie premenné. V tejto súvislosti budeme charakterizovať analýzu ciest a štruktúrne rovnicové modely. Napokon opíšeme základné črty kauzálnej indukcie, ktorá umožňuje nájsť model systému kauzálnych vzťahov, ktorý je v súlade s množinou pozorovaných dát a na ktorom možno realizovať simulačné experimenty a pomocou nich odhadovať účinky zásahov do systému. Problematika analýzy kauzálnych vzťahov je v našej literatúre zatiaľ málo frekventovaná.

2. SKÚMANIE ALTERNATÍVNYCH VYSVETLENÍ ASOCIÁCIE POMOCOU KONTROLY INÝCH PREMENNÝCH

Predpokladajme, že premenná X má kauzálny vplyv na premennú Y . Základným komponentom overovania platnosti tohto predpokladu je skúmanie možných alternatívnych vysvetlení asociácie. Možno ho realizovať prostredníctvom skúmania, či sa asociácia medzi X a Y nestratí, keď odstránime účinky iných premenných na túto asociáciu. Premennú, ktorej vplyv odstránime vo viacrozmernej analýze obyčajne nazývame kontrolovaná (*controlled*) alebo kontrolná (*control*) premenná. Na rozdiel od technických a prírodných vied sa v spoločenskovednej oblasti vo výskume väčšinou využívajú štúdie založené na pozorovaniach. Nemožno jednoducho fixovať hodnoty premenných, ktoré chceme kontrolovať, ako je to bežné v experimentálnych štúdiách. Možno však zoskupiť hodnoty pozorovaní do skupín s rovnakými alebo podobnými hodnotami kontrolovaných premenných. Ak je kontrolovanou premennou napríklad vzdelanie, môžeme rozdeliť pozorovania povedzme do troch skupín: pozorovania na jednotkách s nižším ako stredoškolským vzdelaním, pozorovania na jednotkách so stredoškolským vzdelaním a pozorovania na jednotkách s vysokoškolským vzdelaním. To je štatistická kontrola. V [3] sa na s. 304 - 305 uvádza príklad štúdia asociácie medzi výškou študenta a počtom bodov z matematického testu na základe náhodného výberu študentov jednej americkej strednej školy. Koeficient korelácie vyšiel 0,81, ide teda o silný priamy vzťah, s rastom výšky študenta má počet bodov z testu tendenciu rásť. Môže byť výška študenta príčinou lepších výsledkov z matematiky? Hľadal sa spôsob, ako inak možno túto asociáciu vysvetliť. V náhodnom výbere boli študenti rozličného veku. Starší študenti majú tendenciu byť vyšší a mať aj hlbšie znalosti z matematiky. Účinok veku na asociáciu možno odstrániť prostredníctvom štatistickej

³ O navrhovaní experimentov pozri [8, s. 563 – 645] a [16, s. 319 – 346].

kontroly tak, že analyzujeme asociáciu medzi výškou študenta a počtom bodov z matematického testu zvlášť pre každú vekovú kategóriu študentov. Ukázalo sa, že korelácia medzi výškou študenta a počtom bodov z matematického testu pre študentov z rovnakej vekovej kategórie bola blízka nule pre každú z troch uvažovaných vekových kategórií. To je empirický dôkaz toho, že výška študenta nemá kauzálny účinok na výsledok matematického testu, pretože asociácia zmizla, keď bola veková kategória študentov konštantná. Všeobecne sa študuje vzťah medzi X a Y pre prípady s rovnakou alebo podobnou hodnotou kontrolnej premennej. Tým, že hodnota kontrolnej premennej je konštantná, odstráni sa vplyv tejto premennej na asociáciu medzi X a Y .

Premenná, ktorá sa v štúdiu nemeria, ale ovplyvňuje skúmanú asociáciu, sa niekedy nazýva skrytá premenná (*lurking variable*). Možno súčasne uvažovať o viacerých vysvetľujúcich a kontrolných premenných, ktoré označíme X_1 , X_2 , ... Hovoríme, že asociácia medzi X_1 a Y je zdanlivá (*spurious*), keď obe premenné závisia od tretej premennej X_2 . Zdanlivá asociácia nie je jediná možnosť pri ktorej asociácia zmizne keď kontrolujeme tretiu premennú. Ďalšou možnosťou je kauzálny reťazec (*chain of causation*) v ktorom X_1 ovplyvňuje X_2 , ktoré zasa ovplyvňuje Y . Premenná X_2 sa niekedy nazýva sprostredkovacia premenná (*intervening* alebo *mediator variable*). Niekedy dve premenné nevykazujú asociáciu, pokiaľ sa nekontroluje iná premenná. Táto kontrolná premenná sa najčastejšie nazýva odrušovacia (*suppressor variable*). Často sa účinok vysvetľujúcej premennej na Y mení podľa úrovne inej vysvetľujúcej alebo kontrolnej premennej. Keď sa účinok premennej X_1 na Y mení na rôznych úrovniach X_2 , hovoríme že X_1 a X_2 sú v interakcii. Vtedy má asociácia rozličnú silu a /alebo charakter⁴ na rozličných úrovniach X_2 . Keď dve vysvetľujúce premenné ovplyvňujú Y , ale sú aj navzájom spojené, hovoríme o zmiešaní účinkov (*confounding*). Vtedy je ťažké určiť, ktorá z nich v skutočnosti ovplyvňuje Y , pretože účinok jednej z nich môže byť čiastočne daný jej asociáciou s druhou premennou. Obyčajne pozorujeme iný účinok na Y pre jednu premennú, keď kontrolujeme druhú premennú, ako keď ju nekontrolujeme. Zmiešanie účinkov je v spoločenskovednom výskume celkom bežné. Preto je napríklad ťažké identifikovať príčiny zlepšovania ekonomických výsledkov a pod. (podrobnejšie o typoch vzťahov vo viacrozmernej analýze, pozri v [17, s. 109 - 111]).

Metódy viacrozmernej analýzy ako napríklad viacnásobná regresia umožňujú vykonať štatistickú kontrolu a ohodnotiť vzory asociácie a interakcie bez vykonania čiastkových analýz na rozličných kombináciách úrovní kontrolných premenných. V minulosti sa predpokladalo, že každý regresný model by mal modelovať kauzálne vzťahy medzi premennými. Regresia, ktorá evidentne nebola modelom kauzálnych vzťahov medzi premennými, sa nazývala zdanlivou. Je však užitočné oddeliť aj v regresii asociáciu od kauzality. Asociáciu medzi premennými v regresii možno využiť, aj keď nedáva jednoznačnú odpoveď na otázku o existencii kauzálnych vzťahov medzi nimi. Totiž odhadnutú regresnú rovnicu možno využiť na odhadovanie a na tvorbu predikcií vtedy, keď sú splnené predpoklady regresnej analýzy, vzťah medzi premennými je štatisticky významný a koeficient determinácie ukazuje, že regresia je kvalitná⁵ bez ohľadu na to, či odráža alebo neodráža kauzálne vzťahy

⁴ Zmena priameho vzťahu na nepriamy alebo opačne.

⁵ Podrobnejšie v [16, s. 243].

medzi premennými do tej miery, do akej vzťahy medzi dátami zistené pri odhadovaní zostávajú nezmenené. To sa široko využíva napríklad pri aplikáciách regresnej analýzy v hĺbkovej analýze dát (*data mining*)⁶. Ak je cieľom nájsť najlepší model na predikcie hodnôt nejakej premennej, predikčný model sa formuluje len z dát, bez ohľadu na to, či regresný model odráža, alebo neodráža kauzálne vzťahy medzi premennými. Regresná analýza môže poskytnúť najlepšiu predstavu o existencii a charaktere kauzálnych vzťahov medzi premennými na základe výsledkov realizácie dobre navrhnutého experimentu.⁷

2.1 SIMPSONOV PARADOX

Všimnime si teraz jeden špecifický problém, ktorý komplikuje odhadovanie kauzálnych vzťahov. Je možné, že keď kontrolujeme jednu premennú, každá asociácia v čiastkovej kontingenčnej tabuľke má opačný charakter (podrobnejšie o analýze kategoriálnych dát pozri v [1], [2]). To sa nazýva Simpsonov paradox.

Príklad 1. Máme náhodný výber 800 pacientov, ktorí majú prístup k lieku. 400 pacientov sa rozhodlo liek užívať, 400 pacientov liek neužívalo. Výsledky liečby sú v tabuľke 1. V každom políčku je zľava počet pacientov, v zátvorke je počet uzdravených pacientov a nasleduje percento uzdravených pacientov.

Tabuľka č. 1: Dáta o liečbe

Užívanie lieku	Užíval liek (U)	Neužíval liek (NU)
Pohlavie		
Muži (M)	112 (106), 94,6 %	295 (253), 85,8 %
Ženy (Z)	288 (213), 74 %	105 (74), 70,5 %
Spolu	400 (319), 79,8 %	400 (327), 81,8 %

Zdroj údajov: vlastné výpočty

Podmienenú pravdepodobnosť, že sa muž vylieči (V), keď užíva liek $P(V|U)$ možno odhadnúť pomerom $106/112 = 0,946$. Potom pravdepodobnosť, že sa muž nevylieči (NV), keď užíva liek $P(NV|U) = 1 - P(V|U) = 1 - 0,946 = 0,054$. Podobne $P(V|NU) = 0,858$ a $P(NV|NU) = 0,142$. Odhadnutá šanca, že sa muž vylieči keď užíva liek, je $0,946/0,054 = 17,52$ a odhadnutá šanca že sa vylieči keď neužíva liek, je $0,858/0,142 = 6,04$. Pomer šanci je $17,52/6,04 = 2,9$. Muž má približne 2,9 krát väčšiu šancu, že sa vylieči, keď užíva liek, ako keď ho neužíva. Podobne by sme pre ženy vypočítali, že žena má 1,19 krát väčšiu šancu, že sa vylieči, keď užíva liek, ako keď ho neužíva. Keď však uvažujeme o všetkých pacientoch spolu, nedelíme ich na mužov a ženy, vyjde že pacient má 1,14krát väčšiu šancu, že sa uzdraví, keď liek neužíva, ako keď ho užíva. Zdá sa, že keď poznáme pohlavie pacienta, mali by sme liek odporúčať, keď ho nepoznáme, nemali by sme ho odporúčať. Takýto záver je samozrejme nezmyselný. To, že nepoznáme pohlavie pacienta, nemôže z užitočného lieku urobiť neužitočný.

E. Simpson [14] ako prvý publikoval poznatok o existencii dát, v ktorých charakter asociácie v celom súbore je opačný v každom pod-súbore. Podľa neho sa takáto vlastnosť dát začala nazývať Simpsonov paradox. Aby sme mohli rozhodnúť, či liek pacientovi uškodí alebo pomôže, musíme najprv porozumieť tomu, čo stojí za dátami – príčinnému mechanizmu, ktorý generoval výsledky, ktoré vidíme. Predpokladajme,

⁶ Podrobnejšie o hĺbkovej analýze dát pozri v [15].

⁷ O navrhovaní experimentov v regresii pozri v [16, s. 280 - 281].

že je známe, že estrogén má negatívny vplyv na liečbu, takže je menšia pravdepodobnosť vyliečenia ženy ako muža bez ohľadu na užívanie lieku. Pritom vo výbere pacientov ktorí liek užívali je len $112/400 = 0,28$, t. j. 28 % mužov a 72 % žien. Byť ženou možno považovať za spoločnú príčinu užívania lieku aj jeho menšej účinnosti. Aby sme mohli ohodnotiť účinnosť lieku, musíme porovnávať pacientov rovnakého pohlavia, prípadne štruktúra pacientov podľa pohlavia a užívania lieku musí byť rovnaká. Uzdravenie pacienta závisí od pohlavia aj od užívania lieku. Keď vylúčime vplyv pohlavia, teda pohlavie bude kontrolnou premennou, dostaneme asociáciu medzi uzdravením pacienta a užívaním lieku. Liek by sme mali odporúčať.

Príklad 2. V príklade 1 predpokladajme, že namiesto pohlavia pacientov sa zaznamenával tlak krvi. Predpokladajme, že je známe, že liek ovplyvňuje vyliečenie prostredníctvom zníženia krvného tlaku u tých, ktorí ho užívajú - ale žiaľ, má aj toxický účinok. Dostali sme výsledky uvedené v tabuľke 2.

Tabuľka č. 2: Dáta o liečbe 1

Užívanie lieku	Neužíval liek (NU)	Užíval liek (U)
Krvný tlak		
Nízky krvný tlak (NT)	112 (106), 94,6 %	295 (253), 85,8 %
Vysoký krvný tlak (VT)	288 (213), 74 %	105 (74), 70,5 %
Spolu	400 (319), 79,8 %	400 (327), 81,8 %

Zdroj údajov: vlastné výpočty

V tabuľke 2 sú rovnaké dáta ako v tabuľke 1 s tým rozdielom, že oproti tabuľke 1 sú vymenené kategórie stĺpcov a v riadkoch sú namiesto kategórií pohlavia, kategórie krvného tlaku. Mali by sme v tomto prípade liek odporúčať? Odpoveď opäť vychádza zo spôsobu, akým boli dáta vygenerované. V základnom súbore liek môže zlepšiť mieru vyliečenia kvôli svojmu účinku na krvný tlak, v pod súboroch - skupina ľudí, ktorí mali po liečbe nízky krvný tlak a skupina ľudí, ktorí mali po liečbe vysoký krvný tlak, by sme tento účinok nevideli, videli by sme len toxický účinok lieku. Rovnako ako v príklade 1, cieľom experimentu bolo zistiť celkový účinok užívania lieku na mieru vyliečenia. Tu však zníženie krvného tlaku je jedným z faktorov, ktorými liečebná procedúra ovplyvňuje mieru vyliečenia, a preto nemá zmysel triediť výsledky na základe krvného tlaku. Takže v tomto prípade, vychádzajúc z doplnkových informácií by sme sa mali na rozdiel od príkladu 1 oprieť o výsledky získané za celý základný súbor a liek odporúčať. Dáta v príklade 1 a 2 sú rovnaké, správne rozhodnutie v príklade 1 sa opiera o separované dáta, v druhom príklade o súhrnné dáta. Rozdiel je v doplnkových informáciách, ktoré opisujú príčinný mechanizmus, ktorý generuje dáta. Bez jeho znalosti, len na základe štatistickej analýzy dát by sme nevedeli prijať správne rozhodnutie. Keď skúmame asociáciu medzi premennými s cieľom dozvedieť sa niečo o vzťahoch medzi premennými, nakoniec vždy vychádzame z nejakej základnej predstavy o kauzálnych súvislostiach, ktorá vychádza z nejakých doplnkových informácií.

V experimentálnych štúdiách možno experiment naplánovať tak, aby sme sa vyhli Simpsonovmu paradoxu, generovanému zmiešaním účinkov, samozrejme opäť len na základe vhodných doplnkových informácií (pozri napr. [5]).

3. MODELOVANIE SYSTÉMU KAUZÁLNYCH VZŤAHOV

Opíšeme niektoré možnosti modelovania systému kauzálnych vzťahov a ukážeme, aké predpoklady sú nevyhnutné na odhadovanie ich parametrov v štúdiách založených na pozorovaniach.

3.1 ANALÝZA CIEST

Analýza ciest (*path analysis*) používa na vyjadrenie predpokladov o príčinných vzťahoch medzi premennými, regresné modely, ktoré zahŕňajú príslušné kontrolné premenné. Analytik musí explicitne špecifikovať predpoklady o kauzálnych vzťahoch medzi premennými [4, s. 506]. Začiatočná predstava o kauzálnych vzťahoch medzi premennými často obsahuje taký systém vzťahov, v ktorých vystupujú premenné o ktorých sa predpokladá, že sú dôsledkom iných premenných, ale zároveň ovplyvňujú nejaké ďalšie premenné. Jeden model viacnásobnej regresie nedokáže opísať takýto systém vzťahov, pretože obsahuje len jednu závisle premennú. Analýza ciest využíva sústavu regresných modelov, ktorá dokáže zachytiť všetky predpokladané vzťahy medzi premennými.

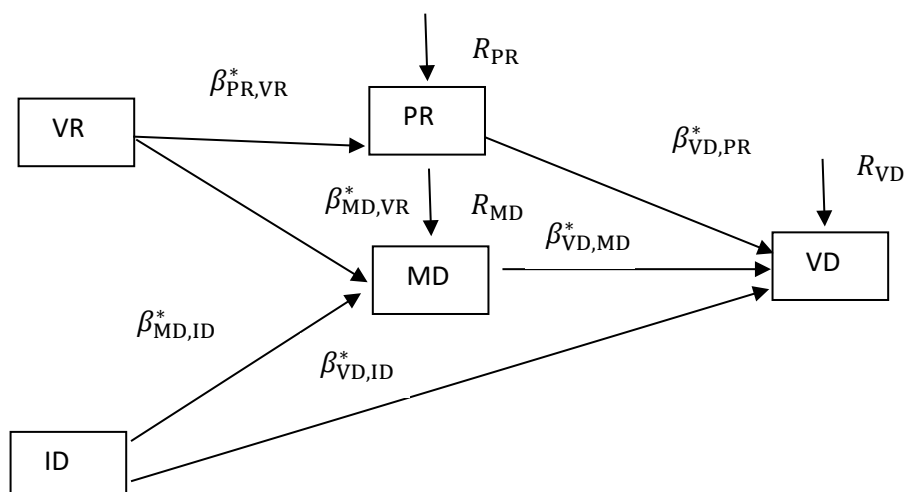
Základným nástrojom na opísanie začiatočnej predstavy o kauzálnych vzťahoch medzi premennými je diagram ciest (*path diagram*). V diagrame ciest je predpokladaný kauzálny vzťah vyjadrený orientovanou hranou, ktorá vychádza z uzla, reprezentujúceho príčinu (nezávisle premenná) a končí sa v uzle, reprezentujúcom následok (závisle premenná). Závisle premenné v príslušných regresných rovniciach korešpondujú s uzlami, v ktorých sa končí aspoň jedna orientovaná hrana. Diagram ciest vyjadri hypotézy o tom, ktoré premenné sú „zodpovedné“ za asociáciu dvoch premenných.

V [4, s. 506 - 509] sa uvádza príklad o analýze vzťahov medzi dosiahnutým vzdelaním dieťaťa VD, jeho motiváciou MD, inteligenciou ID, vzdelaním rodičov VR a príjmami rodičov PR. Predpokladajú sa kauzálne väzby, zobrazené na obrázku 1. Uzly reprezentujú premenné, orientované hrany reprezentujú predpokladané kauzálne väzby. Koeficienty nad hranami (*path coefficients*) sú normované regresné koeficienty (*standardized regression coefficients*). Normovaný regresný koeficient β_i^* pri nezávisle premennej X_i udáva, o koľko svojich smerodajných odchýlok sa zmení stredná hodnota závisle premennej Y , keď sa premenná X_i zväčší o jednu smerodajnú odchýlku Y , keď sú všetky ostatné nezávisle premenné konštantné (kontrolované). Tieto koeficienty sú vhodné na porovnanie relatívnych účinkov viacerých nezávisle premenných na závisle premennú, pretože sa merajú v rovnakých merných jednotkách (v smerodajných odchýlkach závisle premennej Y). Diagram ciest na obrázku č. 1 korešponduje so sústavou troch regresných rovníc:

$$E(VD) = \beta_{VD,PR}^* \cdot PR + \beta_{VD,MD}^* \cdot MD + \beta_{VD,ID}^* \cdot ID$$

$$E(MD) = \beta_{MD,VR}^* \cdot VR + \beta_{MD,ID}^* \cdot ID$$

$$E(PR) = \beta_{PR,VR}^* \cdot VR$$

Obrázok č. 1 Diagram ciest dosiahnutého vzdelania dieťaťa

Zdroj: [4, s. 506 – 509], upravené

Každá zo závisle premenných je v diagrame ciest spojená so svojou reziduálnou premennou (*residual path variable*) [13]. Tieto premenné reprezentujú variabilitu, nevysvetlenú nezávisle premennými v rovnici. Každá z reziduálnych premenných reprezentuje zostávajúci podiel $(1 - R^2)$ nevysvetlenej variability závisle premennej, kde R^2 je koeficient determinácie pre regresnú rovnicu závisle premennej. Normovaný regresný koeficient tejto premennej sa rovná $\sqrt{1 - R^2}$ [4, s. 507]. Jedným zo základných výsledkov analýzy ciest je dekompozícia korelácie medzi dvomi premennými na časti, ktoré súvisia s rozličnými cestami medzi týmito dvomi premennými (podrobnejšie v [4, s. 509 – 510]). Analýza ciest sa všeobecne realizuje v troch krokoch:

1. Predbežná predstava o kauzálnych vzťahoch sa vyjadří pomocou diagramu ciest, bez normovaných regresných koeficientov.
2. Vytvorí sa sústava regresných modelov ktorá slúži na odhadnutie normovaných regresných koeficientov premenných, vrátane reziduálnych premenných.
3. Prostredníctvom kontroly, či sú dáta z náhodného výberu v súlade s modelom, sa model ohodnotí. Prípadné nevýznamné cesty sa zrušia. Modifikovaný model slúži na ďalšiu analýzu. Odhadnú sa normované regresné koeficienty pre tento modifikovaný model.

Na korektné použitie vzťahu na dekompozíciu ciest je v každom z regresných modelov sústavy nevyhnutné splniť predpoklad, že namerané premenné, ktoré sú reprezentované reziduálnou premennou pre závisle premennú v regresnom modeli sú nekorelované s nezávisle premennými v tomto modeli. V príklade by napríklad všetky namerané premenné, ktoré ovplyvňujú dosiahnuté vzdelanie dieťaťa mali byť nekorelované s príjmom rodičov, motiváciou dieťaťa a inteligenciou dieťaťa. To je príliš „silná“ požiadavka. Všeobecne je splnenie tohto predpokladu málo pravdepodobné a prakticky neoveriteľné. Tiež je pravdepodobné, že keby sa do regresných modelov pridali ďalšie premenné, získali by sme odlišný obraz o možných kauzálnych vzťahoch. Nakoniec, ak sú dáta z náhodného výberu aj v súlade s formulovaným diagramom ciest, nedokazuje to pravdivosť kauzálneho systému reprezentovaného týmto diagramom ciest. Pomocou štatistických metód nemožno priamo testovať

predpokladané kauzálne väzby. Ani analýza ciest neumožňuje z asociácie indukovať kauzalitu. Poskytuje len štruktúru na reprezentáciu a odhadovanie predpokladaných kauzálnych účinkov [4, s. 510].

3.2 ŠTRUKTÚRNE ROVNICOVÉ MODELY

Všeobecný štruktúrny rovnicový model (*Structural equation model*) kombinuje prvky analýzy ciest a faktorovej analýzy (podrobnejšie o faktorovej analýze pozri [6]). Model sa nazýva model štruktúry kovariancie (*covariance structure model*), pretože sa snaží vysvetliť rozptyly a korelácie medzi pozorovanými premennými. Toto vysvetlenie má formu kauzálneho modelu, spájajúceho systém faktorov, z ktorých niektoré môžu byť vytvorené vo faktorovej analýze a niektoré môžu byť pozorované premenné. Modely štruktúry kovariancie majú dva komponenty. Prvý je model merania (*measurement model*). Podobá sa na faktorovú analýzu tým, že vytvára množinu nepozorovaných faktorov z pozorovaných premenných. Druhý komponent je štruktúrny rovnicový model. Ten sa podobá na analýzu ciest tým, že špecifikuje regresné modely pre faktory, vytvorené v modeli merania.

Model merania špecifikuje, ako sú pozorované premenné spojené s množinou latentných premenných⁸. Táto časť analýzy sa podobá na faktorovú analýzu ibaže, modelovanie má viac špecifikovaných štruktúru. Model merania priradí každú latentnú premennú a priori k špecifikovanej množine pozorovaných premenných. To sa dosiahne tak, že saturácie⁹ niektorých faktorov sú rovné nule, teda predpokladá sa, že niektoré latentné premenné sú nekorelované s inými premennými. Model merania zohľadňuje fakt, že pozorované premenné môžu byť zaťažené chybami merania a problémami platnosti a spoľahlivosti, a preto nie sú najlepšimi ukazovateľmi konceptov, ktoré nás zaujímajú. Účelom vytvorenia latentných premenných je nájdenie operatívnejších charakteristík, namiesto tých, ktoré je ťažké dobre zmerať, ako sú predsudky, úzkosť, konzervativizmus a podobne.

Štruktúrny rovnicový model používa regresné modely na špecifikáciu kauzálnych vzťahov medzi latentnými premennými. Jedna alebo viaceré z latentných premenných sa špecifikujú ako nezávisle premenné. Latentné závisle premenné môžu v regresnom modeli závisieť od latentných nezávisle premenných ako aj od iných latentných závisle premenných. Na rozdiel od analýzy ciest umožňuje tento prístup aproximáciu modelov s dvojcestnou kauzalitou, v ktorých môžu latentné premenné v regresnom modeli závisieť od ľubovoľných iných latentných premenných.

3.3 KAUZÁLNA INDUKCIA

Na pochopenie metód a modelov kauzálnej indukcie (*causal inference*) je nevyhnutná znalosť základov teórie pravdepodobnosti, štatistiky a teórie grafov. Štruktúrne kauzálne modely (*structural causal models* - SCM) slúžia na opis črt prírody a toho, ako sú spojené s inými črtami, alebo ako príroda priraduje hodnoty premenným, ktoré nás zaujímajú. Formálne je štruktúrny kauzálny model zložený z dvoch množín premenných U a V a z množiny funkcií f , ktoré priradujú každej premennej z V hodnotu, v závislosti od hodnôt iných premenných v modeli. Premenné v množine U sa nazývajú exogénne premenné (*exogenous variables*), čo znamená, že vstupujú do modelu zvonka, nezaujímajú nás, čo je ich príčinou. Premenné v množine

⁸ V štatistike sa nepozorované premenné, ako napríklad faktory, obyčajne nazývajú latentné premenné.

⁹ Korelácia premennej s faktorom sa nazýva saturácia (*loading*) premennej týmto faktorom.

V sú endogénne premenné (*endogenous variables*). Každá endogénna premenná je v orientovanom grafe nasledovníkom (*descendant*) aspoň jednej exogénnej premennej. Exogénne premenné nemôžu byť nasledovníkom žiadnych iných premenných. Nemajú v orientovanom grafe predchodcov (*ancestor*) a sú v ňom reprezentované koreňovými uzlami (*root nodes*). Keď je známa hodnota každej exogénnej premennej, možno pomocou funkcií f determinovať hodnotu každej endogénnej premennej [10, s. 27], [11].

Každý SCM je spojený s grafickým kauzálnym modelom (*graphical causal model*). Ten pozostáva z množiny uzlov, ktoré reprezentujú premenné v množinách U a V a z množiny orientovaných hrán, ktoré ich spájajú. Orientované hrany reprezentujú funkcie z množiny f . Grafický kauzálny model G pre nejaký SCM obsahuje jeden uzol pre každú premennú. Ak funkcia f_X pre premennú X (definuje hodnoty premennej X) obsahuje aj premennú Y (hodnoty X závisia aj od Y), potom v G vedie z Y do X orientovaná hrana. Uvažuje sa len o orientovaných acyklických grafoch G^{10} .

Keď v grafickom kauzálnom modeli, premenná Y je priamym nasledovníkom premennej X , potom X je priamou príčinou Y . Keď Y je nasledovníkom X (nie priamym), potom X je potenciálnou príčinou Y . Definíciu kauzality možno spresniť takto: Premenná X je priamou príčinou (*direct cause*) premennej Y , keď X je vo funkcii, ktorá definuje hodnoty premennej Y . Premenná X je príčinou (*cause*) premennej Y , keď je priamou alebo potenciálnou príčinou Y .

Môže vzniknúť otázka, na čo vlastne používať grafický model, keď obsahuje menej informácií ako SCM. Po prvé, predstavy o kauzálnych vzťahoch medzi premennými obyčajne nie sú kvantitatívne, ako to vyžaduje SCM, ale kvalitatívne, vyjadriteľné v grafickom modeli. Vieme napríklad, že pohlavie človeka je príčinou jeho výšky (muži sú všeobecne vyšší ako ženy) a že výška človeka je jednou z príčin jeho výkonnosti v basketbale, ale numerické ohodnotenie týchto vzťahov nie je jednoduché. Graf kauzálnych vzťahov poskytne čiastočne špecifikovanú verziu SCM. Výhodou grafických kauzálnych modelov je aj to, že umožňujú veľmi efektívne vyjadriť združené rozdelenia pravdepodobnosti. Priameho nasledovníka v modeli nazveme potomkom a priameho predchodcu nazveme rodičom. Pre ľubovoľný acyklický model, združené rozdelenie pravdepodobnosti premenných v modeli je dané súčinom podmienených rozdelení $P(\text{potomok} \mid \text{rodičia})$. Toto pravidlo – pravidlo rozkladu súčinom (*product decomposition rule*) je:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i \mid pa_i) \quad (1)$$

kde pa_i sú hodnoty rodičov premennej X_i [10, s.29].

Proces, ktorý možno použiť v grafickom kauzálnom modeli akejkoľvek zložitosti s cieľom predpovedať závislosti, ktoré zdieľajú všetky súbory dát generované týmto grafom sa nazýva *d*-separácia (*d-separation*) (písmeno *d* znamená *directional* – smerová). Tento proces umožňuje pri ľubovoľnom páre uzlov určiť, či sú *d*-spojené (*d-connected*), to znamená, či medzi nimi existuje cesta, alebo sú *d*-separované (*d-separated*), čo znamená, že medzi nimi neexistuje cesta. Keď povieme, že dva uzly sú *d*-separované, myslíme tým, že premenné, ktoré reprezentujú, sú určite nezávislé.

¹⁰ Graf, ktorý ako podgraf neobsahuje kružnicu (uzavretú postupnosť prepojených vrcholov), sa nazýva acyklický.

Keď povieme, že dva uzly sú d -spojené, myslíme tým, že premenné, ktoré tieto uzly reprezentujú sú možno alebo s najväčšou pravdepodobnosťou závislé [10, s. 46].¹¹

3.3.1 TESTOVANIE MODELU A HĽADANIE KAUZÁLNYCH VZŤAHOV

Keď máme graf G o ktorom sa domnievame, že generuje množinu dát S , d -separácia ukáže, ktoré premenné musia byť podmienené nezávislé¹², v závislosti od ktorých iných premenných. Podmienenú nezávislosť možno testovať pomocou dát. Predpokladajme, že máme zoznam podmienok d -separácie v G a vidíme, že premenné A a B musia byť podmienené nezávislé, v závislosti od C (*independent conditional on C*). Predpokladajme, že odhadneme pravdepodobnosti na základe S a zistíme, že dáta naznačujú, že A a B nie sú podmienené nezávislé v závislosti od C . Potom môžeme zamietnuť G ako možný kauzálny model pre S [10, s. 48 – 49].

Oproti iným metódam testovania adekvátnosti modelu má d -separácia niekoľko výhod. Je neparametrická – nevyžaduje špecifické funkcie na spájanie premenných. Testuje model lokálne, nie globálne. Umožňuje špecifikovať spojenia, v ktorých model nie je v súlade s dátami z výberu a naznačuje, ako možno model opraviť. S využitím d -separácie možno mnoho modelov vylúčiť a dospieť k množine modelov, ktoré nie sú v rozpore s pozorovanými dátami. Význam tohto výsledku spočíva v tom, že dovoľuje nájsť pre kauzálne modely množinu dát, ktorú by mohli generovať. Nielenže možno začať s kauzálnym modelom a generovať množinu dát, ale možno začať s množinou dát a odôvodniť ich pôvod kauzálnym modelom. To je veľmi užitočné, pretože cieľom väčšiny výskumov zameraných na dáta je nájsť model, ktorý ich vysvetľuje.

3.3.2 ÚČINKY ZÁSAHOV

Konečným cieľom mnohých štatistických štúdií je predpovedať účinky zásahov. Keď napríklad zhromažďujeme dáta spojené s povodňami, v konečnom dôsledku nám ide o nájdenie spôsobov, ako možno znížiť ich škodlivé dôsledky a frekvenciu. Tu sa možno oprieť len o dáta zo štúdií založených na pozorovaniach, nie o výsledky realizácie experimentov. Keď kontrolujeme premennú v modeli, fixujeme jej hodnotu. Meníme systém a výsledkom je často zmena hodnôt ďalších premenných. Keď premennú podmienime, nič nezmeníme; iba zúžime náš pohľad na podmnožinu prípadov, v ktorých premenná má hodnotu, ktorá nás zaujíma. To, čo sa potom mení, je naše vnímanie sveta, nie svet samotný. V kauzálnej indukcii sa rozlišuje medzi prípadom, keď premenná X nadobudne hodnotu x prirodzene ($X = x$), a prípadom, keď sa premenná fixuje na hodnote x , zápisom $do(X = x)$. Takže $P(Y = y | X = x)$ je pravdepodobnosť, že $Y = y$, keď sme zistili, že $X = x$, zatiaľ čo $P(Y = y | do(X = x))$ je pravdepodobnosť, že $Y = y$, keď premennú X fixujeme na hodnote x . Pravdepodobnosť $P(Y = y | X = x)$ je pravdepodobnosť rozdelenia pravdepodobnosti Y jednotiek, ktoré majú hodnotu x premennej X . Pravdepodobnosť $P(Y = y | do(X = x))$ je pravdepodobnosť rozdelenia pravdepodobnosti Y jednotiek, keď hodnotu X všetkých jednotiek v základnom súbore fixujeme na úrovni x . Mlčky sa predpokladá, že intervencia nemá žiadne „vedľajšie účinky“, to znamená že priradenie hodnoty x premennej X pre nejakú jednotku nemení priamym spôsobom premenné, ktoré

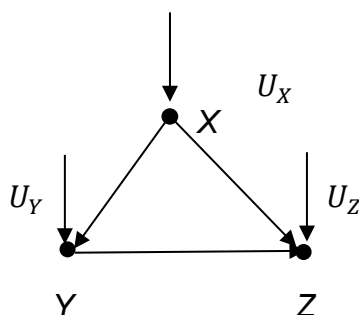
¹¹ Definuje sa aj pojem pravdepodobná závislosť: Premenné Z a Y sú pravdepodobne závislé (*likely dependent*), keď pre niektoré z, y : $P(Z = z | Y = y) \neq P(Z = z)$.

¹² Podmienená závislosť a nezávislosť premenných je definovaná rozlične v rozličných konfiguráciách premenných- reťazce (*chains*), vidlice (*forks*), kolízne schémy (*colliders*).

nasledujú. Napríklad dobrovoľné užívanie lieku môže mať iný vplyv na liečbu jednotlivca ako vynútené užívanie, v rozpore s náboženskými zásadami človeka. Ak sú vedľajšie účinky prítomné, musia byť explicitne zahrnuté do modelu.

Príklad 1 – pokračovanie 1. Máme grafický model, ktorý reprezentuje problém z príkladu 1. Premenná X je pohlavie, Y je užívanie lieku a Z je uzdravenie.

Obrázok č. 2: Grafický kauzálny model, ktorý reprezentuje účinok nového liečiva



Zdroj: [10, s. 55], upravené

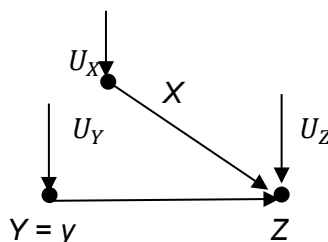
Premenné U_X , U_Y , U_Z na obrázku č. 2, sú exogénne premenné – chybové členy. Ak chceme zistiť, aký účinný je liek v celom základnom súbore, predstavme si hypotetický zásah, ktorým liek podávame všetkým jednotkám základného súboru. Výsledok porovnáme s mierou vyliečenia, keď sa liek nepodáva. Označme prvý zásah ako do ($Y = 1$) a druhý ako do ($Y = 0$). Úlohou je odhadnúť rozdiel:

$$P(Z = 1 | do(Y = 1)) - P(Z = 1 | do(Y = 0))$$

ktorý je známy ako rozdiel v kauzálnom účinku (*causal effect difference*), alebo priemerný kauzálny účinok (*average causal effect - ACE*).

Je známe, že príčinné účinky nemožno odhadnúť zo samotného súboru dát, bez „kauzálneho príbehu“. Samotné dáta z pozorovacej štúdie nestačia na zistenie, či je účinok lieku pozitívny alebo negatívny. Pomocou grafu na obrázku 2 však možno vypočítať veľkosť kauzálneho účinku z dát. Aby sme to dosiahli, simulujeme zásah na obrázku 3. Kauzálny účinok $P(Z = z | do(Y = y))$ sa rovná podmienenej pravdepodobnosti $P_m(Z = z | Y = y)$ z manipulovaného (*manipulated*) modelu na obrázku č. 3.

Obrázok č. 3: Modifikovaný grafický model, ktorý predstavuje zásah do modelu na obrázku 2



Zdroj: [10, s. 56], upravené

Model na obrázku 3 nastavuje užívanie lieku v základnom súbore. Výsledkom je manipulovaná pravdepodobnosť P_m . Kľúč k výpočtu príčinného účinku spočíva v tom, že manipulovaná pravdepodobnosť P_m má dve rovnaké základné vlastnosti s pôvodnou pravdepodobnostnou funkciou v modeli na obrázku 2. Po prvé, marginálna pravdepodobnosť $P(X = x)$ je invariantná k intervencii, pretože proces ktorý determinuje X , nie je ovplyvnený odstránením hrany z X do Y . V príklade to znamená, že proporcie mužov a žien pred intervenciou a po nej zostávajú rovnaké. Po druhé, podmienená pravdepodobnosť $P(Z = z|X = x, Y = y)$ je invariantná, pretože proces, prostredníctvom ktorého Z reaguje na X a Y , $Z = f(y, x, u_z)$, zostáva rovnaký bez ohľadu na to, či sa Y mení spontánne alebo v dôsledku zámernej manipulácie. Môžeme teda napísať dve rovnice invariencie:

$$P_m(Z = z|X = x, Y = y) = P(Z = z|X = x, Y = y) \quad \text{a} \quad P_m(X = x) = P(X = x).$$

Možno využiť aj skutočnosť, že X a Y sú v modifikovanom modeli d -separované, a teda nezávislé. To hovorí, že

$$P_m(X = x|Y = y) = P_m(X = x) = P(X = x).$$

Keď tieto úvahy spojíme, dostaneme:

$$\begin{aligned} P(Z = z|do(Y = y)) &= \\ &= P_m(Z = z|Y = y) = \\ &= \sum_x P_m(Z = z|Y = y, X = x)P_m(X = x|Y = y) = \\ &= \sum_x P_m(Z = z|Y = y, X = x)P_m(X = x) \end{aligned}$$

Nakoniec, keď použijeme vzťahy invariencie, dostaneme vzťah na výpočet kauzálného účinku v termínoch predintervenčných pravdepodobností:

$$P(Z = z|do(Y = y)) = \sum_x P(Z = z|Y = y, X = x)P(X = x) \quad (2)$$

Rovnica (2) sa nazýva vzorec úpravy (*adjustment formula*). Možno podľa neho vypočítať silu asociácie medzi Y a Z pre každú hodnotu x premennej X ($X = 1$ – muž, $X = 0$ – žena), ktorá sa potom spriemeruje cez tieto hodnoty. Tento postup sa označuje ako úprava pre X (*adjusting for X*) alebo kontrola pre X (*controlling for X*).

Pravú stranu v rovnici (2) možno odhadnúť priamo z dát. Pre $Y = 1$ – pacient užíval liek, $Z = 1$ – pacient sa uzdravil, podľa (2) dostaneme

$$P(Z = 1|do(Y = 1)) = P(Z = 1|Y = 1, X = 1)P(X = 1) + P(Z = 1|Y = 1, X = 0)P(X = 0).$$

Po dosadení do predchádzajúceho vzťahu z tabuľky 1, dostaneme

$$P(Z = 1|do(Y = 1)) = 0,946 (112 + 295)/800 + 0,740(288 + 105)/800 = 0,4812775 + 0,363525 = 0,8448025$$

Podobne, pre $Y = 0$ – pacient neužíval liek, $Z = 1$ – pacient sa uzdravil, podľa (2) dostaneme

$$P(Z = 1|do(Y = 0)) = P(Z = 1|Y = 0, X = 1)P(X = 1) + P(Z = 1|Y = 0, X = 0)P(X = 0)$$

Po dosadení do predchádzajúceho vzťahu z tabuľky 1, dostaneme

$$P(Z = 1|do(Y = 0)) = 0,858(112 + 295)/800 + 0,705(288 + 105)/800 = 0,4365075 + 0,34633125 = 0,78283875$$

Keď porovnáme účinok užívania lieku ($Y = 1$) s účinkom jeho neužívania ($Y = 0$), dostaneme

$$ACE = P(Z = 1|do(Y = 1)) - P(Z = 1|do(Y = 0)) = 0,8448025 - 0,78283875 = 0,06196375$$

Z výsledku je zrejмый pozitívny vplyv užívania lieku. Užívanie lieku zväčší pravdepodobnosť uzdravenia v priemere o 0,06196375.

4. ZÁVER

Najsilnejšie indície o existencii a charaktere kauzálnych vzťahov medzi premennými možno získať na základe výsledkov realizácie dobre navrhnutého experimentu, v ktorom možno merať účinok zmeny vstupných premenných na výstupnú premennú. V sociálnoekonomickej oblasti však väčšinou nie je možné navrhnúť a realizovať experimentálne štúdie, pretože niektoré premenné ktoré ovplyvňujú výstupnú premennú nie je možné kontrolovať. Vtedy sa možno oprieť len o štúdie založené na pozorovaniach. Problém týchto štúdií spočíva v tom, že je ťažké oddeliť koreláciu od príčinných súvislostí. Kauzálny vzťah medzi premennými implikuje asociáciu medzi nimi, opačná implikácia však neplatí. Keď sa opierame o dáta zo štúdií založených na pozorovaniach, potom vzťah medzi premennými možno považovať za kauzálny, keď je medzi premennými asociácia, príčina predchádza v čase následok a boli vylúčené alternatívne vysvetlenia asociácie. Splnenie poslednej podmienky je najviac problematické. Nikdy si totiž nemôžeme byť istí, že sme vylúčili všetky iné možné vysvetlenia asociácie a preto nemôžeme nikdy dokázať, že jedna premenná je v kauzálnom vzťahu s druhou.

Pri hľadaní indícií o existencii a charaktere kauzálnych vzťahov medzi premennými je nevyhnutné oprieť sa o nejaké doplnkové informácie umožňujúce opísať predpokladaný príčinný mechanizmus, ktorý generuje dáta. Len na základe štatistickej analýzy dát zo štúdie založenej na pozorovaniach nemožno takéto indície získať. Za každým príčinným záverom musí byť nejaký príčinný predpoklad, ktorý nie je testovateľný na základe dát zo štúdií založených na pozorovaniach [12, s. 99]. V príklade o asociácii medzi počtom mentálne postihnutých osôb pripadajúcich na 10 000 obyvateľov a počtom vydaných licencií na rádiový prijímač v Spojenom kráľovstve sme nakoniec postupovali podobne. Jednoducho sme implicitne konštatovali, že nepoznáme žiadny príčinný mechanizmus, ktorý by mohol generovať dané dáta, a teda napriek asociácii medzi premennými, nejde o kauzálny vzťah.

Ak chceme študovať predpokladané kauzálne vzťahy v modeli systému kauzálnych vzťahov medzi premennými, možno využiť analýzu ciest, ktorá vyjadruje predpoklady o kauzálnych vzťahoch medzi premennými pomocou sústavy regresných modelov. Štruktúrne rovnicové modely kombinujú prvky analýzy ciest a faktorovej analýzy, pričom majú formu kauzálného modelu ktorý sa týka systému faktorov, z ktorých niektoré mohli byť vytvorené vo faktorovej analýze a niektoré môžu byť pozorované premenné. Podobne ako v prípade analýzy ciest, ani pomocou štruktúrnych

rovnícových modelov nemožno priamo testovať predpokladané kauzálne väzby. Aj keď ide o sofistikovanejšiu metódu, ani ona nedokáže indukovať kauzalitu z asociácie. Poskytuje, podobne ako analýza ciest, len štruktúru na reprezentáciu a odhadovanie predpokladaných kauzálnych účinkov.

Za najkomplexnejšiu metodológiu na analýzu kauzálnych vzťahov treba určite považovať kauzálnu indukciu, ktorá umožňuje základnú predstavu o kauzálnych vzťahoch medzi premennými vyjadriť pomocou štruktúrneho kauzálneho modelu, ktorý je spojený s grafickým kauzálnym modelom. V grafickom kauzálnom modeli možno pomocou *d*-separácie špecifikovať spojenia, v ktorých model nie je v súlade s dátami z výberu a naznačuje, ako možno model opraviť. Niektoré modely možno vylúčiť a dospieť k množine modelov, ktoré nie sú v rozpore s pozorovanými dátami. Možno začať aj s množinou dát a odôvodniť ich pôvod kauzálnym modelom. To môže byť často veľmi užitočné. Mimoriadnym prínosom tejto metodológie je možnosť realizovať s modelom simulačné experimenty a odhadovať účinky zásahov do systému. Aplikácia tejto metodológie v praxi však vyžaduje veľmi dobrú znalosť teórie pravdepodobnosti, induktívnych štatistických metód, regresnej analýzy a teórie grafov, spolu s niektorými novými poznatkami, ktoré vznikli v rámci nej. Ani táto metodológia však nedokáže poskytnúť dôkazy o existencii a charaktere kauzálnych vzťahov medzi premennými. Umožňuje len nájsť model kauzálnych vzťahov, ktorý nie je v rozpore s pozorovanými dátami a ponúka nástroje na experimentovanie s týmto modelom a odhadovanie účinkov simulovaných zásahov do systému.

Táto práca bola podporená vedeckou grantovou agentúrou KEGA, v rámci projektu K-20-035-00: Learn Economics: aplikácia e-vzdelávania ako novej formy výučby ekonómie.

LITERATÚRA

- [1] AGRESTI, A.: Analysis of Ordinal Categorical Data. Second Edition. Hoboken: Wiley and Sons, 2010. ISBN 978-0-470-08289-8.
- [2] AGRESTI, A.: Categorical Data Analysis. Third Edition. Hoboken: Wiley and Sons, 2013. ISBN 978-0-470-46363-5.
- [3] AGRESTI, A. – FINLAY, B.: Statistical Methods for the Social Sciences. Fourth edition. Essex: Pearson, 2014. ISBN 978-1-29202-166-9.
- [4] AGRESTI, A.: Statistical Methods for the Social Sciences. Fifth edition. Boston: Pearson, 2018. ISBN 13: 978-0-13-450710-1.
- [5] AMERINGER, S. – SERLIN, R. C. – WARD, S.: Simpson's Paradox and Experimental Research. Dostupné na: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2880329/pdf/nihms-204397.pdf>
- [6] HÄARDLE, W. – HLÁVKA, Z.: Multivariate Statistics: Exercises and Solutions. 2nd edition. Berlin Heidelberg: Springer-Verlag, 2015. ISBN: 978-3-642-36004-6.
- [7] MONTGOMERY, D. C. – PECK, E. A. – VINING, G. G.: Introduction to Linear Regression Analysis. Fifth Edition. Hoboken: J. Wiley and Sons, 2012. ISBN 978-0-470-54281-1.
- [8] MONTGOMERY, D. C.: Statistical Quality Control. A Modern Introduction. Seventh edition. Singapore: J. Wiley and Sons, 2013. ISBN 978-1-118-32257-4.
- [9] MOORE, D. S. – NOTZ, W. I.: Statistics. Concepts and Controversies. New York: W. H. Freeman and Company, 2006. ISBN 978-0-7167-8636-8.
- [10] PEARL, J. – GLYMOUR, M. – JEWELL, N. P.: Causal Inference in Statistics. Chichester: J. Wiley and Sons, 2016. ISBN 9781119186847.

- [11] PEARL, J.: Causality: Models, Reasoning, and Inference. 2nd ed. New York: Cambridge University Press, 2009 (1).
- [12] PEARL, J.: Causal inference in statistics: An overview. In: Statistics Surveys, Vol. 3, 2009 (2), s. 96–146.
- [13] RETHERFORD, R. D. – MINJA KIM CHOE: Statistical Models for Causal Analysis. New York: Wiley and Sons, 1993. ISBN 0-471-55802-8.
- [14] SIMPSON, E.: The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society, 1951, Series B, 13, s. 238 – 241.
- [15] TEREK, M. – HORNÍKOVÁ, A. – LABUDOVA, V.: Hĺbková analýza údajov. Bratislava: Iura Edition, 2010. ISBN 978-80-8078-336-5.
- [16] TEREK, M.: Interpretácia štatistiky a dát. Piate, doplnené vydanie. Košice: Equilibria, 2017. ISBN 978-80-8143-213-2.
- [17] TEREK, M.: Dotazníkové prieskumy a analýzy získaných dát. 1. vydanie. Košice: Equilibria 2019. ISBN 978-80-8143-247-7.

RESUMÉ

Overovanie a odhadovanie predpokladaných kauzálnych vzťahov medzi premennými sa považuje za centrálny problém. Cieľom článku je poskytnúť prehľad o existujúcich postupoch a metódach, ktoré možno v tejto oblasti uplatniť. Sú v ňom opísané ich základné črty. Možnosti ich využívania sú ilustrované na jednoduchých príkladoch.

V článku sú definované pojmy asociácia a kauzalita. Vzťah medzi premennými možno považovať za kauzálny, keď je medzi premennými asociácia, príčina predchádza v čase následok a boli vylúčené alternatívne vysvetlenia asociácie.

Ak predpokladáme, že premenná X má kauzálny vplyv na premennú Y , základným komponentom overovania platnosti tohto predpokladu je skúmanie možných alternatívnych vysvetlení asociácie. Tie možno skúmať pomocou kontroly iných premenných. V štúdiách založených na pozorovaniach možno zoskupiť hodnoty pozorovaní do skupín s rovnakými alebo podobnými hodnotami kontrolovaných premenných – vykonať štatistickú kontrolu. Metódy viacrozmernej analýzy ako napríklad viacnásobná regresia, umožňujú vykonať štatistickú kontrolu a ohodnotiť vzory asociácie a interakcie bez vykonania čiastkových analýz na rozličných kombináciách úrovní kontrolných premenných. Nikdy nemožno dokázať, že jedna premenná je v kauzálnom vzťahu s druhou. Hypotézu o kauzalite možno len vyvrátiť tým, že empirický dôkaz je v rozpore aspoň s jedným z predpokladov kauzality.

Je možné, že keď kontrolujeme jednu premennú, každá asociácia v čiastkovej kontingenčnej tabuľke má opačný charakter. To sa nazýva Simpsonov paradox. V článku sa uvádzajú dva príklady Simpsonovho paradoxu. Dáta v príkladoch sú rovnaké, ale doplnkové informácie opisujúce príčinný mechanizmus, ktorý generuje dáta sú rozdielne, preto sa v rozhodovaní v prvom prípade opierame o separované dáta, v druhom prípade o súhrnné dáta. Bez znalosti príčinného mechanizmu, len na základe štatistickej analýzy dát sa nevieme správne rozhodnúť.

Pri štúdiu predpokladaných kauzálnych vzťahov v modeli systému kauzálnych vzťahov, v ktorom môžu byť premenné, o ktorých sa predpokladá, že sú dôsledkom iných premenných a zároveň ovplyvňujú nejaké ďalšie premenné, možno využiť analýzu ciest, ktorá vyjadruje predpoklady o kauzálnych vzťahoch medzi premennými pomocou sústavy regresných modelov. Štruktúrne rovnicové modely kombinujú prvky analýzy ciest a faktorovej analýzy. Podobne ako v prípade analýzy ciest, ani pomocou štruktúrnych rovnicových modelov nemožno priamo testovať predpokladané kauzálne väzby. Aj keď ide o sofistikovanejšiu metódu, ani ona nedokáže indukovať kauzalitu,

z asociácie. Poskytuje, podobne ako analýza ciest, len štruktúru na reprezentáciu a odhadovanie predpokladaných kauzálnych účinkov.

Za najkomplexnejšiu metodológiu na analýzu kauzálnych vzťahov možno považovať kauzálnu indukciu, ktorá umožňuje základnú predstavu o kauzálnych vzťahoch medzi premennými vyjadriť pomocou štruktúrneho kauzálneho modelu, spojeného s grafickým kauzálnym modelom. V grafickom kauzálnom modeli možno špecifikovať spojenia, v ktorých model nie je v súlade s dátami z výberu a naznačuje, ako možno model opraviť. Mimoriadnym prínosom tejto metodológie je možnosť realizovať s modelom simulačné experimenty a odhadovať účinky zásahov do systému. Ani táto metodológia však nedokáže poskytnúť dôkazy o existencii a charaktere kauzálnych vzťahov medzi premennými. Umožňuje len nájsť model kauzálnych vzťahov, ktorý nie je v rozpore s pozorovanými dátami a ponúka nástroje na experimentovanie s týmto modelom a odhadovanie účinkov simulovaných zásahov do systému.

RESUME

The author considers the verification and estimation of the presumed causal relations between the variables to be a central problem. The aim of the article is to provide an overview of the existing procedures and methods applicable in this field. The article describes their basic features and illustrates the possibilities of their use by simple examples.

The terms association and causality are defined. The relationship between the variables can be considered causal when there is an association between them, the cause precedes the time consequence and alternative explanations of the association were excluded.

Assuming that the variable X has a causal effect on the variable Y , the basic component of validating this assumption is to examine the possible alternative explanations for the association. These can be examined by controlling other variables. In observation-based studies, the observation values can be divided into groups with the same or similar values of the controlled variables - statistical control can be performed. Multivariate analysis methods such as multiple regression enable to perform statistical control and evaluate patterns of association and interaction without performing partial analyzes on various combinations of control variable levels. It can never be proven that one variable is causally related to another. The hypothesis of causality can only be disproved by the fact that empirical evidence contradicts at least one of the causality assumptions.

It is possible that when we check one variable, each association in the partial contingency table has the opposite character. This is called the Simpson's paradox. Two examples of Simpson's paradox are provided in the article. The data in the examples are identical, but the additional information describing the causal mechanism that generates the data is different, so in the first example we rely on separate data, in the second example on the aggregated data. Without knowing the causal mechanism, based only on the statistical data analysis the right decision cannot be made.

In the study of the presumed causal relationships in the model of the system of causal relationships, which may contain variables that are supposed to affect other variables and at the same time also affect some other variables, a path analysis can be used, expressing assumptions about causal relationships between variables using a set of regression models. Structural equation models combine elements of path analysis and factor analysis. Similarly as in case of path analysis, the presumed causal links cannot be directly tested using structural equation models. Although it is a more sophisticated

method, not even that could induce causality from the association. It provides, like path analysis, only a structure for representing and estimating the expected causal effects. Causal inference can be considered the most complex methodology for the analysis of causal relationships, which enables the basic idea of causal relationships between variables to be expressed using a structural causal model, which is associated with a graphical causal model. In the graphical causal model, connections can be specified in which the model does not comply with the sample data and indicates how the model can be corrected. A special benefit of this methodology is the ability to perform simulation experiments with the model and estimate the effects of interventions in the system. However, even this methodology fails to provide proof of the existence and the nature of causal relationships between variables. It only allows to find a model of causal relationships that does not contradict the observed data and offers tools for experimenting with this model and estimating the effects of simulated interventions in the system.

PROFESIJNÝ ŽIVOTOPIS

Prof. Ing. Milan Terek, PhD., od roku 2018 pracuje ako profesor na Vysokej škole manažmentu v Bratislave. Viedie kurzy Úvod do štatistiky, Štatistika, Matematika pre manažérov II, Kvantitatívne metódy pre manažérov a Kvantitatívne metódy vo výskume v oblasti podnikového manažmentu. V rokoch 1977 – 2018 pracoval na Ekonomickej univerzite v Bratislave. Viedol kurzy Štatistika, Štatistické riadenie kvality, Analýza rozhodovania, Hĺbková analýza dát, Výberové skúmanie, Lineárne programovanie, Nelineárne programovanie, Operačný výskum a Systémové modelovanie. Vo výskume sa zameriava na aplikácie štatistických metód v ekonómii a manažmente. Je autorom alebo spoluautorom 6 monografií, 10 vysokoškolských učebníc, 17 skrípt, 75 článkov vo vedeckých a odborných časopisoch a 115 príspevkov na vedeckých konferenciách, publikovaných v zborníkoch.

KONTAKT

mterek@vsm.sk

Boris VAŇO
INFOSTAT – Výskumné demografické centrum

HODNOTENIE KVALITY ADMINISTRATÍVNYCH ÚDAJOV VSTUPUJÚCICH DO ŠTATISTICKÉHO PROCESU V PODMIENKACH SLOVENSKA

QUALITY EVALUATION OF ADMINISTRATIVE DATA ENTERING THE STATISTICAL PROCESS IN CONDITIONS OF SLOVAKIA

ABSTRAKT

Článok poukazuje na význam hodnotenia kvality administratívnych zdrojov údajov pred ich vstupom do štatistického prostredia. Podrobne opisuje schému na hodnotenie kvality AZÚ navrhutej pre potreby štatistickej praxe na Slovensku. Ide o nástroj, ktorý umožňuje Štatistickému úradu SR vykonávať takéto hodnotenie priamo a štandardizovane, v prípade potreby aj opakovane.

ABSTRACT

The article points to the importance of evaluating the quality of administrative data sources before their entry into the statistical environment. It describes in detail the evaluation scheme for evaluating the quality of the ADS designed for the needs of statistical practice in Slovakia. It is a tool allowing the Statistical Office of the SR to perform such assessment directly and in a standardized manner, if necessary also repeatedly.

KLÚČOVÉ SLOVÁ

administratívne zdroje údajov, hodnotenie kvality, schéma na hodnotenie kvality, indikátor kvality, metóda merania

KEY WORDS

administrative data sources, quality evaluation, evaluation scheme, quality indicator, measurement method

1. ÚVOD

Administratívne údaje nie sú primárne určené na účely, na ktoré ich chce využívať štatistický úrad, preto je veľmi dôležité, hodnotiť ich kvalitu, t. j. vhodnosť využitia na štatistické účely. V snahe vyrovnáť sa s výkyvmi v kvalite administratívnych údajov je potrebné, aby štatistický úrad bol schopný hodnotiť ich kvalitu na pravidelnej báze. Štatistický úrad SR (ďalej ŠÚ SR) má na tento účel k dispozícii procedúru, ktorá umožňuje rýchlo, priamo a štandardizovane určiť kvalitu administratívnych zdrojov údajov (ďalej AZÚ) pri ich vstupe do štatistického procesu¹. Takýto prístup je použiteľný pre každý AZÚ, ktorý sa využíva alebo má využívať na štatistické účely.

¹ Hodnotí sa len kvalita AZÚ pred začiatkom ich využívania na štatistické účely. Do úvahy sa neberie ďalšie využívanie AZÚ v rámci štatistického systému, t.j. fáza procesu alebo fáza výstupu.

2. NÁRODNÝ RÁMEC NA HODNOTENIE KVALITY AZÚ URČENÝCH NA ŠTATISTICKÉ ÚČELY

Hodnotenie kvality administratívnych údajov v kontexte ich využívania na štatistické účely je komplexný a odborne náročný proces. Na hodnotenie kvality administratívnych údajov pri ich vstupe do štatistického procesu možno využiť celý rad indikátorov. Pokiaľ má byť hodnotenie objektívne, je potrebné optimalizovať výber jednotlivých indikátorov kvality na základe univerzálnych aj špecifických kritérií. Preto sa prostredníctvom indikátorov kvality vytvárajú tzv. hodnotiace schémy, ktoré sú nástrojmi na komplexné hodnotenie AZÚ pred ich vstupom do štatistického procesu. V hodnotiacich schémach sa využíva rôzna kombinácia indikátorov kvality a ich rôzna váha, ktoré zodpovedajú potrebám a požiadavkám štatistickej praxe v príslušnej krajine a príslušnom období.

Schéma na hodnotenie kvality AZÚ nemá v Európskom štatistickom systéme ani v iných nadnárodných štruktúrach jednotnú podobu a každý štát si ju v konečnom dôsledku vytvára podľa vlastnej potreby a špecifik. V Európskom štatistickom systéme však existujú odporúčania a príklady dobrej praxe a k dispozícii sú aj skúsenosti z krajín, v ktorých takéto hodnotiace schémy existujú už dlhšie obdobie.

Pri tvorbe schémy na hodnotenie kvality AZÚ pre potreby ŠÚ SR sa vychádzalo z odporúčaní Eurostatu [3], [4], [5] aj zo zahraničných skúseností. Postupovalo sa tak, aby sa na jednej strane čo najobjektívnejšie posúdila vhodnosť využitia príslušného AZÚ na štatistické účely a na druhej strane aby sa neznemožnilo jeho využitie pre potreby štatistiky pre menej závažné dôvody. Za základ hodnotiacej schémy pre potreby ŠÚ SR sa zbral teoretický rámec kvality vypracovaný štatistickými úradmi v Holandsku, Taliansku a Švédsku [1], [2], [7], [9]. Pod rámcom kvality treba rozumieť ucelený systém merania kvality s presne zadefinovanou štruktúrou [6], [8].

Na jednoduchšie narábanie s hodnotiacou schémou a na celkovú objektivizáciu hodnotenia kvality AZÚ je potrebné správne a jednotne interpretovať jednotlivé pojmy, ktoré sa nachádzajú v hodnotiacej schéme. Ide hlavne o termíny súvisiace s hierarchickým pohľadom na štatistickú kvalitu administratívnych údajov (hyperdimenzia, dimenzia, indikátor kvality, metóda merania), o termíny spojené s úrovňou aplikácie indikátorov kvality (databáza, objekt, veta, premenná) a o charakteristiky jednotlivých častí administratívnej databázy (objektov, premenných)². V prípade indikátorov kvality sa pri interpretácii venuje pozornosť hlavne kvantitatívnym indikátorom, ktoré sú založené na numerických metódach a ich

² **Hyperdimenzia** predstavuje pohľad na kvalitu na najvyššej hierarchickej úrovni. Každá hyperdimenzia obsahuje najmenej 2 dimenzie.

Dimenzia je druhá najvyššia hierarchická úroveň v kvalitatívnom rámci. Kvalitatívne aspekty obsiahnuté v dimenzii sú výrazne ovplyvnené kontextuálnym pohľadom na kvalitu AZÚ. Každá dimenzia obsahuje najmenej 1 indikátor kvality.

Indikátor kvality charakterizuje špecifickú oblasť, za ktorú sa meria kvalita AZÚ. Pre každý indikátor kvality je k dispozícii najmenej 1 metóda merania.

Metóda merania je kvalitatívna alebo kvantitatívna metóda, ktorá slúži na meranie alebo odhad indikátora kvality.

Štatistická jednotka je jednotka v štatistickom súbore.

Objekt je jednotka v AZÚ. Môže sa líšiť od jednotky v štatistickom súbore, preto je vhodné odlišiť ju aj terminologicky. Ide teda o identifikáciu registrovaných udalostí alebo jednotiek v AZÚ. Z objektov by sa mali dať odvodiť štatistické jednotky.

Veta je skupina údajov, ktoré sa vzťahujú na jeden objekt alebo štatistickú jednotku.

Premenná je konkrétny údaj o objekte alebo štatistickej jednotke, ktorý je súčasťou vety.

interpretácia je podmienená znalosťou týchto metód. Pri kvalitatívnych indikátoroch vyplýva interpretácia viac-menej z názvu indikátora, pričom obvykle ide o otázku na niektorý z aspektov hodnotenia kvality AZÚ.

Pri hodnotení kvality AZÚ využívaných na štatistické účely je veľmi dôležitá spolupráca medzi správcom AZÚ a štatistickým úradom. Preto je pri viacerých indikátoroch kvality potrebná priama komunikácia so správcom AZÚ (väčšinou vo forme poskytnutia určitých informácií).

3. HODNOTIACA SCHÉMA KVALITY AZÚ PRE POTREBY ŠÚ SR

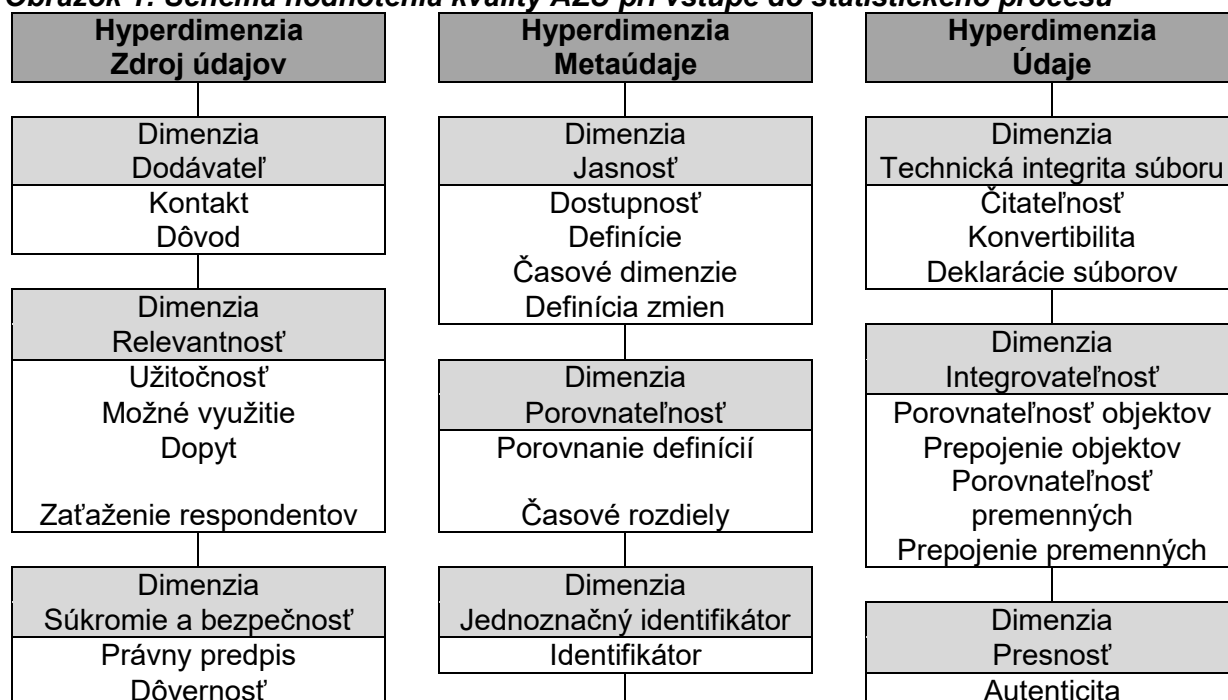
Z teoretického rámca kvality sa v hodnotiacej schéme pre potreby ŠÚ SR využili všetky hyperdimenzie, dimenzie a indikátory kvality, ako aj väčšina metód merania. Hodnotiace kritériá pri jednotlivých indikátoroch kvality aj pri celkovom hodnotení AZÚ sú nastavené v súlade s potrebami štatistickej praxe na Slovensku (s osobitným dôrazom využitie v oblasti sociálnych štatistik a cenzu).

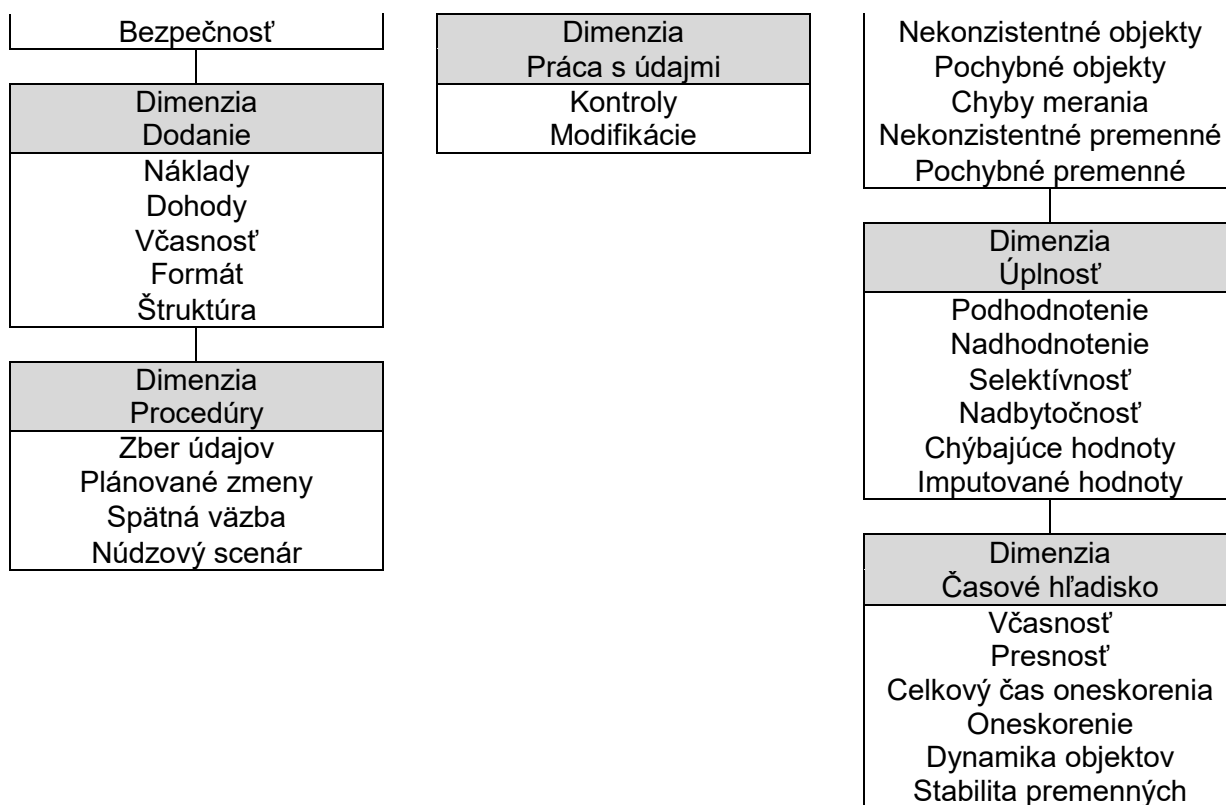
Opis schémy na hodnotenie kvality AZÚ pred ich vstupom do štatistického procesu, navrhutej pre potreby ŠÚ SR obsahuje podrobnú informáciu o všetkých navrhnutých indikátoroch kvality a najčastejšie používaných metódach merania. Je usporiadaný podľa hyperdimenzií a dimenzií [6], [8]. Prehľadné zobrazenie všetkých dimenzií a indikátorov kvality pre všetky tri hyperdimenzie v hodnotiacej schéme sa nachádza na obrázku 1. Prvé dve hyperdimenzie (zdroj údajov a metaúdaje) hodnotia AZÚ ako celok a na hodnotenie využívajú väčšinou kvalitatívne metódy. Tretia hyperdimenzia (údaje) je zameraná na hodnotenie jednotlivých údajov, ktoré sa nachádzajú v AZÚ, pričom na hodnotenie sa využívajú skoro výlučne kvantitatívne metódy.

3.1 Hyperdimenzia zdroj údajov

Hyperdimenzia zdroj údajov obsahuje 5 dimenzií, ktoré pomocou kvalitatívnych metód merania poskytujú základnú charakteristiku príslušného AZÚ. Výnimkou je len dimenzia relevantnosť, v ktorej sa využíva kvantitatívna metóda merania.

Obrázok 1: Schéma hodnotenia kvality AZÚ pri vstupe do štatistického procesu





Zdroj: vlastné

Dimenzia dodávateľ

Indikátory kvality sa v tejto dimenzii zameriavajú na hodnotenie dostupnosti kontaktných informácií o poskytovateľovi AZÚ (indikátor kvality „kontakt“) a hodnotenie dôvodu, pre ktorý je potrebné využívať AZÚ na štatistické účely (indikátor kvality „dôvod“).

Jednotlivé metódy merania pre obidva indikátory kvality sú kvalitatívne. Pri indikátore kvality kontakt ide o otázky na presný názov AZÚ, na kontaktné informácie o správcovi AZÚ (názov a adresa) a na kontaktnú osobu pre ŠÚ SR (meno, telefón, e-mail). Pri indikátore kvality dôvod je jedinou metódou merania otázka na dôvod využívania AZÚ na štatistické účely.

Dimenzia relevantnosť

Táto dimenzia obsahuje 4 indikátory kvality, pričom ku každému indikátoru kvality je priradená jedna metóda merania.

Indikátor kvality „užitočnosť“ hodnotí dôležitosť AZÚ pre ŠÚ SR. Indikátor „možné využitie“ hodnotí možné využitie AZÚ na štatistické účely, t. j. či sú administratívne údaje vhodné na nahradenie, doplnenie alebo kontrolu niektorého štatistického zisťovania. Pomocou indikátora „dopyt“ sa zisťuje, či AZÚ uspokojuje dopyt po informáciách. Indikátor „zaťaženie respondentov“ umožňuje určiť, či využitie AZÚ v štatistickom systéme prispieva k zníženiu zaťaženia respondentov alebo nie. Tento indikátor kvality sa využíva len v prípade, ak AZÚ nahrádza štatistické zisťovanie.

Všetky metódy merania v dimenzii relevantnosť sú kvalitatívne, výnimkou je len indikátor kvality zaťaženie respondentov, ktorý využíva kvantitatívnu metódu merania. Pri kvalitatívnych metódach merania ide o otázku, ktorá hodnotí niektorý z aspektov

relevantnosti AZÚ. Pri indikátore kvality užitočnosť sa otázkou zisťuje, ako dôležitý je AZÚ pre ŠÚ SR. Metóda merania pri indikátore kvality možné využitie zisťuje, či je AZÚ potenciálne vhodný na nahradenie, doplnenie alebo kontrolu niektorého zo štatistických zisťovaní. Otázka pri indikátore spokojnosť zisťuje, či AZÚ uspokojuje dopyt po informáciách.

Metóda merania, ktorou sa zisťuje zaťaženie respondentov, porovnáva zaťaženie respondentov v prípade, že sa pri zisťovaní využívajú, resp. nevyužívajú AZÚ. Na hodnotenie zaťaženia respondentov sa využíva priemerný čas na vyplnenie dotazníka, resp. priemerný čas na získanie jedného údaja.

Dimenzia súkromie a bezpečnosť

V dimenzii súkromie a bezpečnosť sa využívajú 3 indikátory kvality a 5 metód merania zameraných na rôzne aspekty bezpečnosti a ochrany súkromia.

Indikátor kvality „právny predpis“ zisťuje, či existuje právny predpis, ktorý sa týka AZÚ. Indikátor kvality „dôvernosc“ hodnotí ochranu osobných údajov v súvislosti s príslušným AZÚ. Indikátor kvality „bezpečnosť“ hodnotí bezpečnosť pri dodaní AZÚ na ŠÚ SR.

Všetkých 5 metód merania, ktoré sa využívajú v dimenzii súkromie a bezpečnosť, má kvalitatívny charakter. Ide o otázky o existencii právneho predpisu týkajúceho sa príslušného AZÚ (indikátor kvality právny predpis), otázky na využívanie zákona o ochrane osobných údajov a zasielanie správy na Úrad na ochranu osobných údajov (indikátor kvality dôvernosc) a otázky na existenciu špeciálnych dohôd týkajúcich sa bezpečného dodania AZÚ na ŠÚ SR ako aj hardvéru a/alebo softvéru na takéto bezpečné dodanie (indikátor kvality bezpečnosť).

Dimenzia dodanie

Dimenzia dodanie obsahuje 5 indikátorov kvality a 6 metód merania, ktoré sú zamerané na dodanie AZÚ do ŠÚ SR.

Indikátor kvality náklady hodnotí výšku nákladov, ktoré sú spojené s využívaním AZÚ na štatistické účely. Indikátor kvality dohody zisťuje, či je dodávka zdokumentovaná z formálnej stránky. Indikátor kvality včasnosť hodnotí časové hľadisko dodania AZÚ do ŠÚ SR. Indikátor kvality formát zisťuje formát, v ktorom sa AZÚ dostávajú do ŠÚ SR. Indikátor kvality štruktúra hodnotí, či sa administratívne údaje dodávajú do ŠÚ SR v požadovanej štruktúre.

Všetky metódy merania v dimenzii dodanie sú kvalitatívne, ide o otázky týkajúce sa rôznych aspektov dodania AZÚ do ŠÚ SR. Indikátor kvality náklady obsahuje otázku, či sú s dodaním a využívaním AZÚ na ŠÚ SR spojené nejaké náklady. Metóda merania spojená s indikátorom kvality dohody zisťuje, či je dodanie administratívnych údajov do ŠÚ SR zdokumentované. Indikátor kvality včasnosť obsahuje otázku, či sa administratívne údaje dodávajú do ŠÚSR včas, t. j. v súlade s dohodou, a ak dodávka mešká, či je ŠÚ SR o tejto skutočnosti včas informovaný. Metódou merania pri indikátore kvality formát sa zisťuje, či je známy formát, v ktorom sa administratívne údaje dostávajú do ŠÚ SR. Posledná metóda merania, spojená s indikátorom kvality štruktúra, zisťuje, či je štruktúra dodávaných administratívnych údajov v súlade s požiadavkami ŠÚ SR.

Dimenzia procedúry

Dimenzia procedúry obsahuje 4 indikátory kvality a rovnaký počet metód merania. Hodnotenie v rámci tejto procedúry sa zameriava na špecifické, resp. problémové situácie, ktoré by mohli nastať v AZÚ.

Indikátor kvality zber údajov sa týka spôsobu zberu administratívnych údajov, indikátor kvality plánované zmeny zas prípadných zmien v AZÚ. Indikátor kvality spätná väzba zisťuje komunikáciu s dodávateľom AZÚ a indikátor kvality núdzový scenár hodnotí pripravené riešenia v prípade problémov s dodaním AZÚ.

Všetky štyri metódy merania majú podobu otázky a sú teda kvalitatívne. Otázka pri indikátore kvality zber údajov zisťuje, či je ŠÚ SR informovaný o spôsobe zberu administratívnych údajov. Pri indikátore zmeny sa metóda merania zameriava na zistenie, či je ŠÚ SR včas informovaný o prípadných zmenách v AZÚ. Metóda merania pri indikátore kvality spätná väzba umožňuje zistiť, či má ŠÚ SR možnosť kontaktovať dodávateľa AZÚ v prípade problémov. Metóda merania pri indikátore kvality núdzový scenár zisťuje, či existuje núdzový scenár pre prípad problémov s dodaním AZÚ. Ide predovšetkým o problémy s termínom dodania a kvalitou dodaných údajov.

3.2 Hyperdimenzia metaúdaje

Hyperdimenzia metaúdaje hodnotí kvalitu AZÚ v zmysle definícií, porovnateľnosti, identifikácie a kontrolnej činnosti zo strany poskytovateľa AZÚ. Jasnosť definícií a kompletnosť metainformácií patria medzi hlavné kvalitatívne aspekty v tejto hyperdimenzii. Hyperdimenzia metaúdaje obsahuje 4 dimenzie a 12 metód merania.

Dimenzia jasnosť

Dimenzia jasnosť hodnotí jasnosť údajov z rôznych aspektov. Na tento účel využíva 4 indikátory kvality a rovnaký počet metód merania. Indikátor kvality dostupnosť zisťuje existenciu metaúdajov a či boli dodávateľom ŠÚ SR aj poskytnuté. Indikátor kvality definície sa zameriava na definovanie údajov v AZÚ prostredníctvom metaúdajov. Ide o definície zisťovaných jednotiek, klasifikačných premenných a početností. Indikátor kvality časové dimenzie hodnotí administratívne údaje z časového hľadiska. Indikátor kvality definícia zmien sa zameriava na zmeny a úpravy v definíciách jednotlivých údajov v AZÚ.

Indikátor kvality dostupnosť využíva ako metódu merania otázku, či sú k dispozícii metaúdaje a či boli poskytnuté ŠÚ SR spolu s AZÚ. Metódou merania pri indikátore kvality definície je otázka, či sú údaje nachádzajúce sa v AZÚ jasne definované v metaúdajoch. Otázka, ktorá prislúcha k indikátoru kvality časové dimenzie, zisťuje, či je časové obdobie, na ktoré sa vzťahujú údaje v AZÚ, jasne opísané. Indikátor kvality zmeny v definíciách využíva ako metódu merania otázku, či nastali zmeny v definíciách, a ak áno, či boli jasne a včas odkomunikované.

Dimenzia porovnateľnosť

Dimenzia porovnateľnosť hodnotí porovnateľnosť AZÚ so štandardmi, ktoré sa používajú v ŠÚ SR. V tejto dimenzii sú dva indikátory kvality a dve metódy merania. Indikátor kvality porovnanie definícií hodnotí porovnateľnosť definícií používaných v AZÚ a v ŠÚ SR. Ako už bolo spomenuté pri predchádzajúcej dimenzii, ide o definície

zisťovaných jednotiek, klasifikačných premenných, ako aj početností. Indikátor kvality „Časové rozdiely“ hodnotí porovnateľnosť časových referencií používaných v AZÚ a v ŠÚ SR.

Obidva indikátory kvality sú kvalitatívne, ako metódu merania využívajú otázky, ktoré zisťujú, či sú definície v AZÚ a v ŠÚ SR porovnateľné (indikátor kvality porovnanie definícií) a či sú časové referencie používané v AZÚ a na ŠÚ SR porovnateľné (indikátor kvality časové rozdiely).

Dimenzia jednoznačný identifikátor

Dimenzia jednoznačný identifikátor má jeden indikátor kvality a dve metódy merania. Kvalitatívny indikátor identifikátor hodnotí existenciu a porovnateľnosť jednoznačného identifikátora v AZÚ. Na hodnotenie využíva dve otázky, pomocou ktorých zisťuje, či v AZÚ existuje jednoznačný identifikátor a či je porovnateľný s identifikátormi používanými v ŠÚ SR.

Dimenzia práca s údajmi

Táto dimenzia hodnotí prácu správcu AZÚ s údajmi v zdroji. Ide o všetky vykonané zmeny a kontrolnú činnosť správcu. Dimenzia obsahuje 2 indikátory kvality a 4 kvalitatívne metódy merania. Na hodnotenie v rámci tejto dimenzie sú potrebné podklady od poskytovateľa AZÚ.

Indikátor kvality kontroly hodnotí kontrolnú činnosť správcu AZÚ. Ide o kontrolu zisťovaných jednotiek, kontrolu premenných, kontrolu kombinácie premenných, kontrolu extrémnych hodnôt. Indikátor kvality modifikácie hodnotí zasahovanie do AZÚ zo strany správcu. Za modifikácie sa považuje editovanie, imputácie a dosadzovanie difoltných hodnôt. Hodnotenie kvality sa zameriava na oboznámenosť s modifikáciou údajov, označenie modifikovaných údajov a oboznámenosť s použitím difoltných hodnôt. Metódy merania pri oboch indikátoroch kvality majú podobu otázok. V prípade indikátora kvality kontroly sa otázkami zisťuje, či správca kontroloval údaje v AZÚ a či sa pri kontrole špeciálne zameril na výskyt extrémnych hodnôt. Pri indikátore kvality modifikácie sa zisťuje, či správca modifikoval údaje v AZÚ a či sú modifikované hodnoty v AZÚ aj vyznačené.

3.3 Hyperdimenzia údaje

Hyperdimenzia údaje sa venuje rôznym aspektom kvality údajov, ktoré sa nachádzajú v príslušnom AZÚ. Je zo všetkých hyperdimenzií najrozsiahlejšia, obsahuje 5 dimenzií, 25 indikátorov kvality a 39 metód merania. Na hodnotenie kvality sa v tejto hyperdimenzii využívajú takmer výlučne kvantitatívne metódy merania.

Dimenzia technická integrita súboru

Dimenzia technická integrita súboru obsahuje 3 indikátory kvality a 3 metódy merania. Indikátor kvality čitateľnosť hodnotí dostupnosť súboru a všetkých údajov v súbore. Problémom z hľadiska čitateľnosti môže byť napríklad, keď je súbor v neznámom formáte, keď je súbor poškodený alebo keď nie je možné súbor dekódovať. Za čitateľný sa považuje taký súbor, ktorý má čitateľné všetky vety. AZÚ sa považuje za čitateľný, ak sú čitateľné všetky jeho súbory. Metóda merania, ktorá v hodnotiacej schéme prislúcha indikátoru kvality čitateľnosť, hodnotí podiel nečitateľných súborov z celkového počtu dodaných súborov (obvykle v percentách).

Alternatívne možno nečitateľnosť hodnotiť aj kapacitným vyjadrením (napríklad v MB) ako rozsah poškodených súborov k celkovému rozsahu dodaných súborov.

Indikátor kvality konvertibilita hodnotí možnosť previesť súbor do formátu, ktorý je štandardom v ŠÚ SR. Problémom konvertibility môže byť napr. poškodenie súboru pri dekódovaní alebo chybné údaje v súbore po konverzii. Počas konverzie môžu vzniknúť chyby, v dôsledku ktorých nie je možné súbor otvoriť, alebo v súbore, ktorý je možné otvoriť, sú chybné niektoré údaje. Metóda merania, ktorú využíva indikátor kvality konvertibilita, udáva podiel objektov s chybou pri dekódovaní alebo s chybnými údajmi po konverzii z celkového počtu objektov v AZÚ.

Indikátor kvality deklarácie súborov hodnotí súlad údajov v AZÚ s informáciami v metaúdajoch. Rozdiely medzi údajmi v súbore a metaúdajmi môžu byť v názvoch a formátoch, v obsahu premenných a príznakov, v kategóriách definovaných pre kvalitatívne premenné a v rozsahu pre kvantitatívne premenné. Problémom môže byť napr. chýbajúci popis metaúdajov alebo údaje v súbore s nesprávnou hlavičkou. Zásadnou požiadavkou je, aby boli metaúdaje k dispozícii. Metóda merania použitá v hodnotiacej schéme vyjadruje podiel premenných v AZÚ, ktoré sa líšia od špecifikácie uvedenej v časti metaúdaje.

Dimenzia integrovateľnosť

V dimenzii integrovateľnosť sa nachádzajú 4 indikátory kvality a 9 metód merania. Všetky metódy merania v tejto dimenzii sú kvantitatívne.

Indikátor kvality porovnateľnosť objektov hodnotí podobnosť objektov v AZÚ s objektmi používanými v ŠÚ SR. Na hodnotenie kvality sa využíva len v prípade, ak sa AZÚ používa na nahradenie, doplnenie alebo kontrolu existujúceho štatistického zisťovania. V AZÚ môžu byť identické, zodpovedajúce a neintegrovateľné objekty. Identické administratívne objekty majú rovnakú definíciu a jednotku ako štatistické objekty. Zodpovedajúce administratívne objekty nie sú identické so štatistickými objektmi, ale je možné ich konvertovať. Neintegrovateľné administratívne objekty nie sú zhodné so štatistickými objektmi a nie je ich možné ani integrovať. Metódy merania pri indikátore kvality porovnateľnosť objektov využívajú jednotlivé typy administratívnych objektov. To znamená, že určujú podiel identických, zodpovedajúcich alebo neporovnateľných administratívnych objektov na celkovom počte relevantných objektov v AZÚ.

Indikátor kvality prepojenie objektov hodnotí možnosť prepojenia (párovania) objektov v AZÚ a štatistickom súbore. Identické administratívne objekty sú zhodné so štatistickými objektmi, preto ich prepojenie je bezproblémové. Zodpovedajúce administratívne objekty možno prepojiť so štatistickými objektmi až po konverzii. Nespojené administratívne objekty nie je možné spojiť so žiadnym štatistickým objektom. Za východisko pri prepájaní sa vždy považujú objekty v štatistickom súbore. Metódy merania pri tomto indikátore kvality vyjadrujú podiel identických priradených objektov, zodpovedajúcich priradených objektov, nepriradených objektov alebo nepriradených agregovaných objektov na celkovom počte objektov v AZÚ.

Indikátor kvality porovnateľnosť premenných hodnotí blízkosť hodnôt premenných v AZÚ s hodnotami premenných v iných zdrojoch používaných v ŠÚ SR. Metódou merania je podiel objektov s identickými hodnotami premenných v oboch zdrojoch z celkového počtu objektov v oboch zdrojoch (štatistickom aj administratívnom).

Indikátor kvality prepojenie premenných hodnotí užitočnosť prepájacích premenných (napr. identifikačných čísel) pre objekty v AZÚ. Problémom môžu byť objekty bez prepájacích premenných alebo objekty s prepájacími premennými, ktoré sa líšia od

tých, ktoré používa štatistický úrad. Ako metódu merania možno využiť podiel objektov s neprepojenými premennými, podiel objektov s prepojenými premennými, ktoré sú rozdielne od premenných používaných v štatistickom úrade, alebo podiel objektov s korektne konvertovateľnými prepojenými premennými z celkového počtu objektov s prepojenými premennými v AZÚ.

Dimenzia presnosť

Presnosti, ako kritériu kvality sa v hodnotiacej schéme venuje veľká pozornosť. V tejto dimenzii sa nachádza 6 indikátorov kvality a 7 kvantitatívnych metód merania. Ako alternatíva k jednej z kvantitatívnych metód merania sa môže využiť 5 kvalitatívnych metód, ktoré majú podobu otázok.

Indikátor kvality autenticita je zameraný na hodnotenie legitimacy objektov v AZÚ. Autenticita zahŕňa syntaktickú správnosť použitých identifikátorov (ak existujú) a zhodu objektov v AZÚ s príslušnými reálnymi objektmi. Problémom je objekt s nesprávnym identifikátorom (syntaktická chyba), resp. syntakticky správny ale nesprávne priradený identifikátor. Pokiaľ existuje v zdroji viac identifikátorov (napr. identifikátor domácnosti a identifikátor osoby), treba skontrolovať správnosť každého z nich. Na identifikáciu objektov s nesprávne priradeným identifikátorom je potrebné porovnanie s referenčným zoznamom. Objekty vyhodnotené ako problematické z hľadiska tohto indikátora kvality sa označujú ako neautentické. Na hodnotenie ako metóda merania slúži podiel objektov so syntakticky nesprávnym identifikátorom z celkového počtu objektov v AZÚ alebo podiel objektov, ktoré v súbore obsahujú protichodné informácie v porovnaní s informáciami v referenčnom zozname pre tieto objekty (z celkového počtu objektov v AZÚ).

Indikátor kvality nekonzistentné objekty je zameraný na celkovú konzistentnosť objektov v AZÚ. Nekonzistentné objekty sú zapojené do nelogických vzťahov s inými objektmi. Indikátor sa používa, len ak sú v zdrojovom súbore viacnásobné objekty. Problémom je napr., ak je jedna osoba priradená k viacerým domácnostiam alebo lokálna obchodná jednotka k viac ako jednému podniku. Metódou merania sa v tejto dimenzii zisťuje podiel objektov zapojených do nelogických vzťahov s inými objektmi z celkového počtu objektov v AZÚ.

Indikátor kvality pochybné objekty sa zameriava na existenciu nedôveryhodných objektov. Ide o objekty, ktoré majú nepravdepodobný, avšak nie jednoznačne nesprávny vzťah s inými objektmi. Identifikácia pochybných objektov si vyžaduje dobrú znalosť súboru. Indikátor má zmysel len v prípade, že súbor obsahuje viacnásobné objekty. Za príklad pochybného objektu možno označiť domácnosť, ktorá má 25 členov, t. j. 25 osôb (objektov) je priradených k tej istej domácnosti, alebo manželov vo veku nad 100 rokov. Metódou merania sa zisťuje podiel objektov zapojených do nepravdepodobných (nie však nevyhnutne nesprávnych) vzťahov s inými objektmi z celkového počtu objektov v AZÚ.

Indikátor kvality chyby merania hodnotí korektnosť hodnôt premenných v AZÚ. Meria odchýlku aktuálne nameraných údajov od ideálnych bezchybných meraní (zisťovaní), resp. rozsah, v akom nameraná hodnota zodpovedá hodnote, ktorá mala byť nameraná. Existuje viacero dôvodov, pre ktoré nemá zistená hodnota potrebnú validitu. Môže to byť napríklad chyba pri hlásení (evidované objekty majú záujem byť evidované určitým spôsobom) alebo chyba pri spracovaní (administratívna prax alebo

spôsob spracovania môže mať za následok skreslené údaje). Keďže všetky dôvody súvisia s procesom získavania alebo spracovania údajov, je nevyhnutné získať potrebné informácie na výpočet tohto indikátora od poskytovateľa AZÚ. Pri tomto indikátore kvality sa v hodnotiacej schéme nachádza jedna kvantitatívna metóda merania, ktorá udáva podiel neoznačených chybných hodnôt zo všetkých hodnôt v AZÚ. Táto metóda je použiteľná len v prípade, že chybné hodnoty sú v AZÚ vyznačené. Ak poskytovateľ AZÚ chybné hodnoty nevyznačil, je potrebné ho kontaktovať a hodnotenie získať na základe niekoľkých otázok, ktoré sú zamerané na uplatnenie stanoveného dizajnu v procese zberu údajov, kontrolu údajov počas fázy reportingu, používanie štandardov pre príslušné premenné, kontrolný proces pre vstup údajov a kontroly počas spracovania alebo ukladania údajov.

Indikátor kvality nekonzistentné premenné hodnotí premenné z hľadiska konzistentnosti hodnôt v prípade kombinácií premenných v AZÚ, t. j. do akej miery sú údaje v AZÚ vnútorne konzistentné. Hodnotenie konzistentnosti si vyžaduje dobrú znalosť údajov v súbore. Príkladom nekonzistentnosti je napr. kombinácia vek 10 rokov a rodinný stav rozvedený. Inak povedané, ide o všetky kombinácie hodnôt premenných, ktoré odporujú niektorej definícii. Metóda merania pre tento indikátor kvality udáva podiel objektov, ktoré obsahujú nelogickú kombináciu hodnôt premenných z celkového počtu objektov v AZÚ.

Indikátor kvality pochybné premenné zisťuje existenciu nepravdepodobných kombinácií hodnôt premenných objektu. Ide o kombinácie, ktoré sú podozrivé, nie však nevyhnutne nesprávne. Obvykle ide o kombinácie hodnôt, ktoré výrazne prevyšujú zvyčajné hodnoty, napríklad podnik, ktorý má 10-krát vyšší obrat, ako je obvyklé pri danom type podnikov, alebo matka s 15 deťmi. Posúdenie pochybnosti hodnôt vyžaduje dobrú znalosť údajov v AZÚ. Metóda merania v tomto prípade udáva podiel objektov, ktoré obsahujú nepravdepodobné kombinácie hodnôt premenných z celkového počtu objektov v AZÚ.

Dimenzia úplnosť

Dimenzia úplnosť hodnotí rôzne stránky úplnosti súboru, nie len v zmysle chýbajúcich alebo nadbytočných hodnôt, ale aj ich podhodnotenia alebo nadhodnotenia ako aj metód riešenia úplnosti. Obsahuje 6 dimenzií a 7 metód merania. Všetky metódy merania sú kvantitatívne.

Indikátor kvality podhodnotenie hodnotí, či v AZÚ chýbajú objekty, ktoré sa nachádzajú v referenčnom zozname. Referenčným zoznamom, s ktorým sa porovnávajú zdrojové údaje, je cieľový zoznam, ktorý berie do úvahy ŠÚ SR. Príkladom chýbajúcich objektov sú napríklad osoby, ktoré boli sčítané v sčítaní obyvateľov a spĺňajú podmienku na zaradenie do registra obyvateľov, ale v čase sčítania sa v registri nenachádzajú. Metódou merania je podiel objektov z referenčného zoznamu, ktoré chýbajú v AZÚ (z celkového počtu objektov v referenčnom zozname).

Indikátor kvality nadhodnotenie hodnotí existenciu iných ako cieľových objektov v AZÚ. Ide o také objekty, ktoré v príslušnom referenčnom období nepatria do cieľového súboru štatistického úradu. Napríklad, keď sa v registri obyvateľov v čase sčítania nachádzajú osoby, ktoré neboli sčítané. Metódou merania je podiel objektov v AZÚ, ktoré sa nenachádzajú v referenčnom zozname (z celkového počtu objektov).

Indikátor kvality selektívnosť sa týka štatistického pokrytia a reprezentatívnosti objektov v AZÚ. Indikátor je zameraný hlavne na objekty, ktoré nechýbajú náhodne. Na posúdenie selektivity je potrebná informácia o celom súbore. Ako referenčný súbor sa použije cieľový zoznam príslušnej štatistiky. Selektivita sa v hodnotiacej schéme meria pomocou indikátora reprezentatívnosti (R – indikátor) alebo prostredníctvom tabuľkového diagramu.

Indikátor kvality nadbytočnosť hodnotí viacnásobnú evidenciu identických objektov v AZÚ, a to buď na základe duplicity identifikátorov, ktoré majú byť evidované len raz alebo na základe duplicitnej hodnoty pre vybrané premenné, ktoré slúžia na identifikáciu. Za duplicitné objekty sa považujú také objekty, ktoré majú rovnaký identifikátor alebo ktoré majú rovnakú hodnotu pre vybrané premenné. Na meranie duplicitnosti sa v hodnotiacej schéme využíva podiel duplicitných objektov z celkového počtu objektov v AZÚ.

Indikátor kvality chýbajúce hodnoty sa týka chýbajúcich hodnôt premenných v AZÚ. Môže ísť o jednotlivú premennú alebo o kombináciu premenných. Hodnotenie si vyžaduje dobrú znalosť súboru. Na meranie sa využíva podiel objektov s chýbajúcou hodnotou pre príslušnú premennú (zo všetkých objektov v AZÚ).

Indikátor kvality imputované hodnoty sa týka existencie doplnených hodnôt v AZÚ zo strany poskytovateľa. Štatistický úrad vie identifikovať imputované hodnoty len v prípade, ak ich poskytovateľ označí. Pokiaľ sú imputované hodnoty vyznačené v súbore, hodnotenie nevyžaduje ďalšiu spoluprácu zo strany poskytovateľa AZÚ. Pokiaľ nie sú imputované hodnoty vyznačené, musí ich poskytovateľ pred hodnotením vyznačiť. Hodnotenie sa robí na základe podielu imputovaných hodnôt, ktoré pripadajú na jednu premennú.

Dimenzia časové hľadisko

Posledná dimenzia v hyperdimenzii dáta je zameraná na časové hľadisko dodania AZÚ do ŠÚ SR a na zmeny, ktoré sa udiali v porovnaní s predchádzajúcim dodaným súborom (v prípade, že sa AZÚ nedodáva do ŠÚ SR prvý raz). Ide o dôležitú stránku hodnotenia kvality, pretože oneskorené doručenie môže ovplyvniť použiteľnosť údajov. Dimenzia časové hľadisko obsahuje 6 indikátorov kvality a 8 metód merania. Všetky metódy merania v tejto dimenzii sú kvantitatívne.

Indikátor kvality včasnosť hodnotí časový rozdiel medzi koncom referenčného obdobia a okamihom prevzatia AZÚ štatistickým úradom. Metódou merania je časový rozdiel (obvykle v dňoch), ktorý sa určí ako rozdiel medzi dátumom doručenia AZÚ do ŠÚ SR a dátumom konca referenčného obdobia, na ktoré sa viažu údaje v AZÚ.

Indikátor kvality presnosť hodnotí časový rozdiel medzi dátumom doručenia AZÚ do ŠÚ SR a dohodnutým dátumom doručenia. Metódou merania je časový rozdiel (obvykle vyjadrený v dňoch), ktorý sa pri tomto indikátore kvality počíta ako rozdiel medzi skutočným a dohodnutým dátumom doručenia AZÚ do ŠÚ SR.

Indikátor kvality celková doba oneskorenia hodnotí celkový časový rozdiel medzi koncom referenčného obdobia a okamihom, keď štatistický úrad potvrdil, že súbor sa môže začať využívať. Na rozdiel od predchádzajúcich indikátorov kvality, tento indikátor berie do úvahy aj čas potrebný na vyhodnotenie dodaných údajov. Metódou

merania je pri tomto indikátore celkový časový rozdiel (v dňoch), ktorý sa vypočíta ako rozdiel medzi dátumom, keď ŠÚ SR deklaruje, že AZÚ môže začať využívať, a dátumom konca referenčného obdobia, na ktoré sa viažu údaje v AZÚ.

Indikátor kvality oneskorenie hodnotí rýchlosť, s akou sa prípadné zmeny v AZÚ zachytia v evidencii. Týka sa to zmien v zložení súboru aj zmien v hodnotách premenných pre objekty v súbore. Súbory s veľkým oneskorením poskytujú štatistickému úradu viac alebo menej zastarané údaje. Využitie tohto indikátora vyžaduje podklady od poskytovateľa AZÚ. Metódou merania je časový rozdiel v dňoch, ktorý sa vypočíta ako rozdiel medzi dátumom zachytenia zmeny v AZÚ zo strany poskytovateľa a dátumom, keď zmena nastala.

Indikátor kvality dynamika objektov hodnotí identifikáciu zmien v súbore v priebehu času. Ide o zmeny, ktoré sú spôsobené vznikom a zánikom objektov. Objekty, ktoré zostávajú v súbore, sa nazývajú pretrvávajúce objekty. Adekvátne zaraďovanie nových objektov a vyradovanie zaniknutých objektov je nevyhnutný predpoklad na kvalitnú aktualizáciu súboru.

Pre objekty v súbore platí nasledujúca klasifikácia (t je aktuálne obdobie):

nové objekty (t) = objekty existujúce v čase t avšak nie v čase $t-1$

pretrvávajúce objekty (t) = objekty existujúce v čase t a zároveň v čase $t-1$

zaniknuté objekty (t) = objekty existujúce v čase $t-1$ avšak nie v čase t

všetky objekty (t) = nové objekty (t) + pretrvávajúce objekty (t)

všetky objekty ($t-1$) = pretrvávajúce objekty (t) + zaniknuté objekty (t)

Dynamiku objektov je možné hodnotiť pomocou troch metód merania. Ide o podiel nových, zaniknutých alebo pretrvávajúcich objektov z celkového počtu objektov v AZÚ.

Indikátor kvality stabilita premenných hodnotí zmeny premenných v čase. Má zmysel len pre pretrvávajúce objekty. Pre niektoré premenné je nevyhnutné, aby sa v čase nemenili (napr. klasifikačné kódy a pod.). Kvalita sa v tomto prípade hodnotí pomocou podielu objektov so zmenenými hodnotami z celkového počtu objektov v AZÚ. Do hodnotenia sa zaraďujú len pretrvávajúce objekty a stabilné premenné, teda tie, ktoré by sa v čase nemali meniť.

4. POUŽITIE HODNOTIACEJ SCHÉMY

Opísaná schéma na hodnotenie kvality AZÚ vstupujúcich do štatistického systému obsahuje 51 indikátorov kvality a 74 metód merania. Ku každému indikátoru kvality je priradená najmenej jedna metóda merania. Pre niektoré indikátory kvality je v hodnotiacej schéme k dispozícii viac metód merania, ktoré však môžu byť navrhnuté alternatívne. Pokiaľ sú metódy merania navrhnuté alternatívne, vyberie sa len jedna z metód a tá sa použije na hodnotenie. Výber jednej z alternatívnych metód je plne v kompetencii hodnotiteľa. Kritériom pri výbere môže byť napríklad dostupnosť údajov alebo zložitosť výpočtu. Pokiaľ nie sú metódy merania navrhnuté alternatívne, môže sa pre príslušný indikátor kvality použiť viac metód merania.

Na základe kvalitatívnej alebo kvantitatívnej hodnoty, ktorá sa získa pomocou príslušnej metódy merania (odpoveď na otázku alebo výpočet ukazovateľa), sa indikátoru kvality pridelia body. Bodové hodnotenie je nastavené tak, aby čo najviac zohľadňovalo realitu, priority a požiadavky štatistickej praxe na Slovensku. Pokiaľ je pre indikátor kvality navrhnutých viac metód merania (a nie sú navrhnuté alternatívne),

tak sa indikátoru priradí hodnotenie za každú použitú metódu merania. Takto sa postupuje pri každom indikátore kvality a celkové hodnotenie pre AZÚ sa získa ako súčet hodnotení za všetky indikátory kvality.

Nie je nutné využiť na hodnotenie AZÚ všetky indikátory kvality. To znamená, že ľubovoľný indikátor kvality možno vynechať, či už z dôvodu nepotrebnosti, alebo z dôvodu neexistencie požadovaných údajov (môže nastať hlavne v prípade, keď poskytovateľ administratívnej databázy nedodá potrebné informácie na hodnotenie). Viaceré indikátory kvality sú bezpredmetné v prípade, že v štatistickom systéme neexistuje štatistické zisťovanie s rovnakým alebo podobným zameraním ako hodnotený AZÚ. Pokiaľ však hodnotený AZÚ vstupuje do štatistického systému s cieľom nahradiť, doplniť alebo kontrolovať niektoré existujúce štatistické zisťovanie, sú relevantné prakticky všetky indikátory kvality, teda aj tie, ktoré pri hodnotení berú do úvahy viaceré zdroje údajov.

V praxi sa na hodnotenie kvality AZÚ využíva automatizovaná verzia hodnotiacej schémy [8]. Ide o používateľský softvér vytvorený ako nadstavba nad MS Excel, v programovacom jazyku VBA (Visual Basic for Applications). Pri hodnotení sa do vyznačených polí jednoducho zadávajú požadované vstupy (odpovede na otázky alebo údaje na výpočet kvantitatívnych metód merania) alebo sa metóda merania označí ako nehodnotená. Tieto vstupné informácie majú podobu externých vstupov a pripravujú sa mimo hodnotiacej schémy. Príprava vstupných informácií vyžaduje dobrú znalosť hodnoteného súboru a v niektorých prípadoch treba na získanie požadovaných údajov použiť aj zložitejšie procedúry.

Na základe zadaných vstupov program priradí metóde merania kvalitatívnu alebo numerickú hodnotu a na jej základe priradí následne aj počet bodov indikátoru kvality. Po prejdení všetkých indikátorov kvality sa prechádza na záverečné zhrnutie, ktoré poskytuje informáciu o celkovom hodnotení AZÚ v percentách. Ide o podiel dosiahnutých bodov z celkového počtu možných bodov. Okrem celkového skóre zhrnutie obsahuje aj prehľadné hodnotenie jednotlivých indikátorov kvality v podobe podielu dosiahnutých bodov pre príslušný indikátor z maximálne možného počtu bodov pre príslušný indikátor. Do celkového hodnotenia sa započítajú len použité indikátory kvality, vynechané indikátory sa nezarátavajú a v hodnotení sú zobrazené ako nehodnotené.

5. ZÁVER

Hodnotiaca schéma kvality AZÚ je určená na hodnotenie AZÚ, ktoré vstupujú do štatistického systému. Umožňuje priamo, cielene a štandardizovane hodnotiť všetky AZÚ, ktoré chce ŠÚ SR využívať na štatistické účely. Napĺňa sa tým potreba hodnotiť kvalitu AZÚ pred vstupom do štatistického prostredia, pretože AZÚ nie sú primárne určené na využitie pre potreby štatistickej praxe.

Dôležitým aspektom hodnotenia kvality AZÚ je fakt, že v prípade niektorých AZÚ prebieha hodnotenie kvality opakovane a výsledky hodnotenia sa premietajú do komunikácie s jednotlivými správcami AZÚ s cieľom zlepšiť využitie AZÚ na štatistické účely.

Opísaný spôsob hodnotenia kvality AZÚ bol do značnej miery implementovaný aj na účely SODB 2021, a to tak v etape prípravy, ako aj v etape spracovania údajov.

Napomohol najmä pri konečnom rozhodnutí o kvalite AZÚ a jeho následnom využití na SODB 2021, čo sa premietlo do zákona o sčítaní. Ďalej napomohol pri určení, ktoré údaje sa budú zbierať a ktoré sa plne prevezmú z AZÚ. Pomohol tiež vymedziť, ktoré AZÚ, resp. premenné z AZÚ sa použijú na kontrolu premenných v SODB 2021.

LITERATÚRA

- [1] Cerroni, F., Di Bella, G., Galié, L.: Evaluating administrative data quality as input of the statistical production process. In: *Rivista di Statistica Ufficiale*, 2014, č. 1-2, s. 117-146.
- [2] Daas, P., Ossen, S., Vis-Visschers, R., Arends-Tóth, J.: Checklist for the quality evaluation of administrative data sources. The Hague/Heerlen, Statistics Netherlands, 2009, 18 s.
- [3] Eurostat. 2003. Quality assessment of administrative data for statistical purposes. Luxembourg. 22 s.
- [4] Eurostat. 2011. Kódex postupov pre európsku štatistiku. Luxembourg. 8 s.
- [5] Eurostat. 2014. ESS handbook for quality reports, 2014 edition. Luxembourg, Office for official publications of the European communities, 139 s. ISBN 978-92-79-45487-5.
- [6] Ivančíková, Ľ., Vaňo, B.: Teoreticko-metodologické aspekty hodnotenia administratívnych zdrojov údajov využívaných na štatistické účely. *Slovenská štatistika a demografia*, 2021, č.(3), s. 57-71.
- [7] Laitila, T., Wallgren, A., Wallgren, B.: Quality assessment of administrative data. Stockholm, Statistics Sweden, 2011. 33 s. ISSN 1653-7149.
- [8] Vaňo, B.: Národný rámec na hodnotenie kvality administratívnych údajov určených na štatistické účely. Bratislava, INFOSTAT, 2017. 39 s.
- [9] Zhang, L.: Topics of statistical theory for register-based statistics. *Statistica Neederlantica*, 2012, 66, č. 1, s. 41-66.

RESUMÉ

Administratívne údaje nie sú primárne určené na účely, na ktoré ich chce využívať štatistický úrad. Pred vstupom do štatistického prostredia je preto veľmi dôležité, hodnotiť ich kvalitu, t. j. vhodnosť využitia na štatistické účely. Štatistický úrad SR má na tento účel k dispozícii procedúru, ktorá umožňuje rýchlo, priamo a štandardizovane určiť kvalitu administratívnych zdrojov údajov.

Hodnotenie kvality administratívnych údajov v kontexte ich využívania na štatistické účely je komplexný a odborne náročný proces. Na hodnotenie kvality administratívnych údajov možno využiť celý rad kvalitatívnych aj kvantitatívnych indikátorov. Pokiaľ má byť hodnotenie objektívne, je potrebné optimalizovať výber jednotlivých indikátorov kvality na základe univerzálnych aj špecifických kritérií. Preto sa prostredníctvom indikátorov kvality vytvárajú tzv. hodnotiace schémy, ktoré sú nástrojom na komplexné hodnotenie administratívnych zdrojov údajov.

Z hierarchického hľadiska sú hodnotiace schémy usporiadané do hyperdimenzií a dimenzií. V každej dimenzii sa nachádzajú indikátory kvality (kvalitatívne alebo kvantitatívne) a ku každému indikátoru kvality prináleží najmenej jedna metóda merania.

Hodnotiacia schéma navrhnutá pre Štatistický úrad SR využíva na hodnotenie kvality administratívnych zdrojov údajov vstupujúcich do štatistického procesu, 3 hyperdimenzie, 14 dimenzií, 52 indikátorov kvality a 74 metód merania. Prvé dve hyperdimenzie (zdroj údajov a metaúdaje) hodnotia administratívny zdroj údajov ako celok a na hodnotenie využívajú väčšinou kvalitatívne metódy. Tretia hyperdimenzia

(údaje) je zameraná na hodnotenie jednotlivých údajov, ktoré sa nachádzajú v administratívnom zdroji údajov, pričom na hodnotenie sa využívajú skoro výlučne kvantitatívne metódy.

V praxi sa na hodnotenie kvality AZÚ využíva automatizovaná verzia hodnotiacej schémy. Ide o používateľský softvér, ktorý zjednodušuje a zrýchľuje hodnotenie a eliminuje formálne chyby.

RESUME

Administrative data are not primarily used for the purposes for which the Statistical office intends to use them. Therefore it is very important to evaluate their quality, i.e. suitability for statistical purposes, before entering the statistical environment. For this purpose, the Statistical Office of the SR has a procedure at its disposal, enabling fast, direct and standard determination of the quality of administrative data sources.

Evaluating the quality of administrative data in the context of their use for statistical purposes is a complex and professionally demanding process. A number of qualitative and quantitative indicators can be used to assess the quality of administrative data. If the evaluation is to be objective, it is necessary to optimally determine the selection of individual quality indicators on the basis of universal and specific criteria. Therefore, through quality indicators, the so-called evaluation schemes are created, serving as a tool for a comprehensive evaluation of administrative data sources.

From a hierarchical perspective, evaluation schemes are arranged in hyperdimensions and dimensions. Each dimension contains quality indicators (qualitative and quantitative) and there is at least one measurement method for each quality indicator. The evaluation scheme, designed for the Statistical Office of the SR, uses for the quality evaluation, administrative data sources entering the statistical process and it has 3 hyperdimensions, 14 dimensions, 52 quality indicators and 74 measurement methods. The first two hyperdimensions (Data Source and Metadata) evaluate the administrative data source as a whole and mostly use qualitative methods for evaluation. The third hyperdimension (Data) focuses on the evaluation of individual data contained in an administrative data source, using almost exclusively quantitative evaluation methods.

In practice, an automated version of the evaluation scheme is used for the quality assessment of the ADS. It is a user software simplifying and accelerating the evaluation and eliminating the formal errors.

PROFESIJNÝ ŽIVOTOPIS

Ing. Boris Vaňo vyštudoval Vysokú školu ekonomickú v Bratislave, následne absolvoval postgraduálne štúdium z demografie na Karlovej Univerzite v Prahe. Od roku 1980 pracuje v Inštitúte informatiky a štatistiky ako výskumný pracovník v oblasti demografie. V rokoch 2000 – 2014 bol vedúcim Výskumného demografického centra, v období rokov 2006 – 2010 podpredsedom Slovenskej štatistickej a demografickej spoločnosti pre demografiu. Špecializuje sa na hodnotenie populačného vývoja, demografické prognózy a populačnú politiku.

KONTAKT

vano@infostat.sk

Recenzia publikácie/Review of Publication

Michal Páleš, František Slaninka:
TEÓRIA RIZIKA V POISTENÍ
Riešené príklady v jazyku R a Maxima

Michal Páleš, František Slaninka
RISK THEORY IN INSURANCE
Solved Examples in language R and Maxima

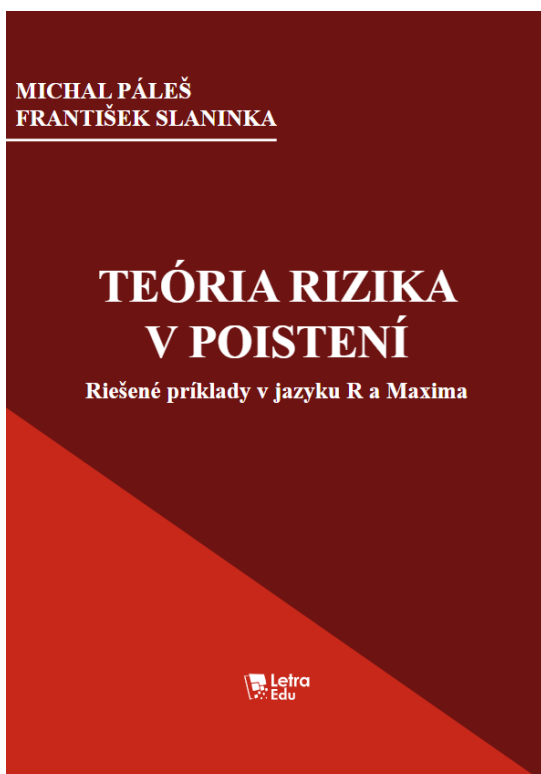
Prvé vydanie.

Bratislava: Letra Edu, 2021, 203 s.

ISBN 978-80-89962-80-8 (print)

ISBN 978-80-89962-81-5 (online)

Vysokoškolský učebný text *Teória rizika v poistení, riešené príklady v jazyku R a Maxima* bol vydaný v tlačenej i v online verzii vo vydavateľstve Letra Edu, Bratislava v roku 2021. Autormi učebného textu sú členovia Katedry matematiky a aktuárstva Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave. Fakulta hospodárskej informatiky okrem iného poskytuje od roku 1994 vzdelanie aj v študijnom programe aktuárstvo (predtým poistná matematika a štatistika) na druhom stupni vysokoškolského štúdia. V tomto študijnom programe sa vyučujú i predmety teória rizika v poistení I a teória rizika v poistení II. Dlhé roky predmet učila, rozvíjala a garantovala doc. RNDr. Galina Horáková, CSc. Prednášky a cvičenia k týmto predmetom boli v roku 2015 základom na napísanie učebnice *Teória rizika v poistení*, ktorej autormi sú Galina Horáková, Michal Páleš a František Slaninka. Cieľom aktuálne recenzovaného učebného textu je využitie jazyka R a Maxima na konfrontovanie riešených príkladov zo spomínanej učebnice z roku 2015.



Učebný text má 203 strán a pozostáva zo šiestich kapitol, dodatku a príloh, ktoré zahŕňajú prehľad základných rozdelení pravdepodobnosti spolu s ich charakteristikami.

Publikácia v úvode čitateľa uvádza do problematiky riadenia rizík v poisťovni a predstavuje i prierez základných informácií a postupu inštalácie využívaných softvérov (jazyk R a Maxima).

Učebný text obsahovo úzko korešponduje s publikáciou z roku 2015 a jeho hlavným cieľom je riešiť príklady, ktoré sú v učebnici „manuálne“ vypočítané, predovšetkým využitím metódy simulácií Monte Carlo a R funkcie *replicate* na určenie hodnôt náhodnej premennej rozdelenia celkovej škody

a následne v aplikáciách kolektívneho modelu rizika, resp. poukázaním na využitie R

knižnice s názvom *actuar*. Systém Maxima je využitý najmä v problematike optimalizácie zaist'ovacích ochrán a tiež tam, kde je potrebné použiť výlučne matematický softvér. Takto je možné numerické postupy teórie rizika viacmenej plnohodnotne nahradiť technikami, ktoré využívajú benefity poskytované modernou výpočtovou technikou.

Jednotlivé názvy kapitol učebného textu korešpondujú s názvami hlavných kapitol v učebnici z roku 2015, pričom tieto kapitoly uvádzajú len kľúčové teoretické aspekty. Teóriu k riešenej problematike autori vynechali s odkazom na pôvodnú publikáciu, kde sa interakcia s ňou pri štúdiu danej oblasti predpokladá i odporúča.

V prvej kapitole s názvom *Stochastický prístup k riadeniu rizík* sa praktické výpočty týkajú analýzy, modelovania a riadenia rizika vo financiách a v poisťovníctve a sú založené na štatistických metódach, keď na základe pozorovaných údajov a pomocou vhodných štatistických nástrojov môžeme uskutočniť závery o pravdepodobnostnom správaní príslušných javov.

Druhá kapitola *Individuálny model rizika* sa venuje analýzam v životnom poistení.

Tretia kapitola *Kolektívny model rizika* uvádza metódy na určenie rozdelenia celkovej škody. Tu sa opäť ukazuje, že možnosti využitia pokročilých softvérových prostredí a výpočtovej techniky sa stali doslova neoceniteľnými v oblasti riadenia rizík v poisťovni. Tieto postupy sa však môžu zaujímavo využiť i v oblasti finančnictva, zdravotníctva alebo aj v oblasti marketingu (a inde).

Štvrtá kapitola *Netto poistné pre rôzne formy poistenia* v príkladoch opäť využíva simulácie Monte Carlo a funkciu *replicate*, simulované hodnoty už stačí len upraviť v zmysle príslušnej formy poistenia.

Jazyk R poskytuje na analýzu zaistenia, ktorou sa zaoberá piata kapitola *Zaistenie*, niekoľko užitočných procedúr, najmä simulácie Monte Carlo, ako aj špecifické knižnice (napr. *ReinsureR* a *ReIns*). Pri optimalizačných kritériách založených na stanovení viazaných extrémov je podľa publikácie vhodné využiť systém Maxima.

Posledná, šiesta kapitola *Kolektívny model rizika na dlhšie časové periódy* sa venuje stochastickým procesom, ktoré sú súčasťou analýz teórie krachu a ukazuje efektivitu využitia simulácií, a teda nevyhnutnosť využitia vhodného softvéru vzhľadom na náročnosť samotného výpočtu.

Koncepciu učebného textu ocenia najmä vysokoškolskí študenti študijného programu aktuárstvo, rovnako aj doktorandi, ktorí sa venujú danej oblasti. Takto prakticky orientovaná publikácia má všetky predpoklady stať sa i vítaným doplnkom knižnice každého záujemcu o problematiku riadenia rizík v poisťovni, rovnako študenta, ako aj profesionála.

Mgr. Andrea Kaderová, PhD.

Autorka pôsobí na Katedre matematiky a aktuárstva Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave.

PRIPRAVUJEME/COMING SOON

Ľudmila IVANČÍKOVÁ

IMPLEMENTÁCIA RÁMCU HODNOTENIA KVALITY PRI SČÍTANÍ OBYVATEĽOV,
DOMOV A BYTOV 2021

IMPLEMENTATION OF THE QUALITY ASSESSMENT FRAMEWORK FOR THE
2021 POPULATION AND HOUSING CENSUS

Branislav ŠPROCHA

POKRAČUJÚCA TRANSFORMÁCIA SOBÁŠNOSTI SLOBODNÝCH NA
SLOVENSKU V LONGITUDINÁLNOH POHĽADE[1]

CONTINUING TRANSFORMATION OF NUPTIALITY IN SLOVAKIA IN A
LONGITUDINAL PERSPECTIVE

Michal HRNČIAR

SEKTOROVO ŠPECIFICKÁ NEZAMESTNANOSŤ

SECTOR - SPECIFIC UNEMPLOYMENT

* * *

**ONLINE VERZIA ČÍSLA 2/2022 SLOVENSKEJ ŠTATISTIKY A DEMOGRAFIE JE
VEREJNE DOSTUPNÁ** na internetovej stránke ssad.statistics.sk od **15. apríla 2022**

**THE ONLINE VERSION OF THE JOURNAL SLOVAK STATISTICS AND
DEMOGRAPHY No 2 (2022) IS PUBLICLY BE AVAILABLE** at
the website ssad.statistics.sk from **APRIL 15, 2022**

INFORMÁCIE PRE PRISPIEVATEĽOV

Príspevky prijímame v slovenskom, v českom a v anglickom jazyku. Musia rešpektovať odborné zameranie časopisu a jeho vedecký charakter. Zaslaný príspevok nesmie byť v recenznom konaní v inom časopise, ani uverejnený v odbornej a inej tlači.

Príspevky zasielajte v elektronickej forme vo formáte MS Word alebo Open Office, typ písma Arial, veľkosť 12, riadkovanie 1. Nad titulkom treba uviesť meno autora a jeho pracovisko.

Súčasťou príspevku je abstrakt (základný popis cieľa a spôsobu spracovania faktov v rozsahu do 100 slov), kľúčové slová (maximálne 5), resumé (stručné zhrnutie obsahu článku s dôrazom na jeho prínos a najvýznamnejšie závery v rozsahu do 500 slov), profesijný životopis (v rozsahu do 120 slov) a kontakt (e-mailová adresa autora). Názov článku, abstrakt, kľúčové slová a resumé poskytne autor aj v anglickom jazyku. Zoznam použitej literatúry v abecednom poradí s úplnými bibliografickými údajmi sa uvádza na konci článku. Odkazy na literatúru sa uvádzajú v texte číslami v hranatých zátvorkách. Poznámky s poradovým číslom sú umiestnené pod čiarou na príslušnej strane textu, ku ktorému sa vzťahujú. Podrobnejšie pokyny nájdete autori na ssad.statistics.sk.

Maximálny rozsah vedeckých článkov je 15 normostrán, informatívnych článkov 6 normostrán, recenzie, rozhovory a informácie publikujeme v rozsahu maximálne 3 normostrany. Tabuľky, mapy, grafy a obrázky musia mať názov a uvedený zdroj údajov; odporúčame, aby kopírovali šírku textu. Skratky sa používajú len minimálne, pri prvom použití je potrebné skratku v zátvorke rozpísať. Redakcia zabezpečuje jazykovú úpravu textu.

Príspevky sú recenzované. Oponentské konanie je obojstranne anonymné. Konečné rozhodnutie o publikovaní článku vydáva redakčná rada.

Redakcia si vyhradzuje právo zverejniť články schválené redakčnou radou v tlačenej a elektronickej podobe na ssad.statistics.sk.

INFORMATION FOR AUTHORS

Articles are accepted in Slovak, Czech and English languages and must comply with the journal's professional specialisation and scientific nature as well. The submitted articles should not be reviewed by another journal and should not have already been published in any specialised or other press.

Please submit your articles in electronic form, in MS Word or Open Office format, Arial font, size 12 and typed in single spacing. The author's name and workplace should be indicated above the title.

Articles should contain an abstract (general description of the objective and the processing methods used up to 100 words), key words (max. 5), resume (brief summary of the article's content emphasizing its contribution and the most important conclusions up to 500 words), curriculum vitae of the author (no more than 120 words) and the author's contact (e-mail address). The author should submit the article's title, abstract, key words and resume in English language. List of the literature used with full bibliographic data should be given in alphabetical order at the end of an article. Bibliographic citations should be given in square brackets. References are indicated by numbers in a text in square brackets. Footnotes should be numbered in the order of the corresponding page of a text. Authors can find more details at the website ssad.statistics.sk.

Maximum scope of a scientific article is up to 15 standard pages, informative articles should be up to 6 standard pages in length, reviews, discussions and information not more than 3 standard pages. Tables, maps, graphs and pictures should have a title and the data source indicated, it is also advised to copy the width of a text. Abbreviations should be used only rarely and should be appropriately explained in parentheses when first used. Language text revisions are provided by the editorial office.

Articles are reviewed. The opponent procedure is mutually anonymous. The final decision on the article's publication is made by the editorial board.

The editorial office reserves the right to publish articles approved by the editorial board in printed and electronic form at the website ssad.statistics.sk.

SLOVENSKÁ ŠTATISTIKA A DEMOGRAFIA

je jediný recenzovaný vedecký časopis so zameraním na prezentáciu moderných štatistických a demografických metód a postupov. Propagujeme miesto a význam slovenskej štatistiky v Európskom štatistickom systéme, spoluprácu Eurostatu a národných štatistických úradov pri harmonizácii zisťovaní a multidimenzionálny rozmer štatistiky. Podporujeme rozvoj štatistickej teórie a jej prepojenie s praxou. Naším cieľom je prispievať k využiteľnosti štatistických výstupov v rôznych oblastiach a k zvyšovaniu ich kvality a efektivity.

Publikujeme analytické články, prognózy, názory, diskusné príspevky, recenzie, rozhovory, informácie a oznamy z rôznych oblastí štatistiky (národné účty, produkčné štatistiky, sociálne štatistiky, štatistika životného prostredia a pod.) a demografie (demografická štatistika, teoreticko-metodologické východiská demografie, historická demografia a pod.), vrátane sčítania obyvateľov, domov a bytov ako neodmysliteľnej súčasti demografickej štatistiky.

Vydáva:

Štatistický úrad SR

Identifikačné číslo vydavateľa:

IČO 00166197

Vychádza:

Štyrikrát ročne

Dátum vydania:

15. apríl 2022

Tlač:

Reprografické stredisko
Štatistického úradu SR

Predplatné:

20 € (na rok)
5 € (za jeden výtlačok)

Objednávky prijíma:

Informačný servis
Štatistického úradu SR
Tel.: +4212/502 36 339
+4212/502 36 335
E-mail: info@statistics.sk

SLOVAK STATISTICS AND DEMOGRAPHY

is the only scientific reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures. Our aim is to promote the position and importance of Slovak statistics in the European Statistical System, cooperation between the Eurostat and the national statistical offices in the field of survey harmonisation and the multidimensional character of statistics as well. We support the development of statistical theory and its connection with practice. We aim to contribute to the utility of statistical outputs in various fields and to the improvement of quality and efficiency.

We publish analytic articles, prognoses, views, discussion contributions, reviews, discussions, information and announcements from various statistical fields (national accounts, production statistics, social statistics, environmental statistics etc.) and demography (demographic statistics, theoretical and methodological bases of demography, historical demography etc.) including the population and housing census as an essential part of demographic statistics.

Issued by:

Statistical Office of the SR

Company registration number:

00166197

Published:

Four times a year

Date of issue:

15th April 2022

Press:

Reprographic centre of the
Statistical Office of the SR

Subscription:

€20 (per year)
€5 (for one copy)

Orders are to be addressed to:

Information Service of the
Statistical Office of the SR
Tel.: +4212/502 36 339
+4212/502 36 335
E-mail: info@statistics.sk