

Štatistický úrad Slovenskej republiky
The Statistical Office of the Slovak Republic
Slovenská štatistická a demografická spoločnosť
The Slovak Statistical and Demographic Society

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

vedecký časopis/scientific journal

4/2024
ročník 34



ŠTATISTICKÝ
ÚRAD
SLOVENSKEJ
REPUBLIKY



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

SLOVENSKÁ ŠTATISTIKA A DEMOGRAFIA

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

Zahranční 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. Individual copies of the journal are available to readers in electronic form at the websites ssad.statistics.sk and slovak.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

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
VZDELANIE, EKONOMICKÁ AKTIVITA A ZDRAVOTNÝ STAV OSÔB
V MARGINALIZOVANÝCH RÓMSKYCH KOMUNITÁCH NA SLOVENSKU PODĽA
VÝSLEDKOV ZISŤOVANIA EU-SILC MRK 2020
EDUCATION, ECONOMIC ACTIVITY AND HEALTH STATUS OF PERSONS
IN MARGINALIZED ROMA COMMUNITIES IN SLOVAKIA ACCORDING TO THE
RESULTS OF THE EU-SILC MRC 2020 SURVEY

Roman PAVELKA 21
IMPUTACE CHYBĚJÍCÍCH DAT POMOCÍ BAYESOVSKÉHO MODELOVÁNÍ
MISSING DATA IMPUTATION USING BAYESIAN DATA MODELLING

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

Emília ČIČVÁKOVÁ 42
ŽENY NA TRHU PRÁCE NA SLOVENSKU A VO VEREJNOM ŽIVOTE
V POROVNANÍ S KRAJINAMI EURÓPSKEJ ÚNIE
WOMEN IN THE LABOUR MARKET IN SLOVAKIA AND IN PUBLIC LIFE
COMPARED TO THE COUNTRIES OF THE EUROPEAN UNION
Informatívny článok/Informative article

Jasmína STAUDER 51
DRUHÝ ROČNÍK EURÓPSKEJ SÚŤAŽE V ŠTATISTIKE: SLOVENSKÍ
STREDOŠKOLÁCI OPĚŤ EXCELUJÚ
SECOND EDITION OF THE EUROPEAN STATISTICS COMPETITION: SLOVAK
HIGH SCHOOL STUDENTS EXCEL ONCE AGAIN
Informácia/Information

Iveta STANKOVIČOVÁ, Silvia HUDECOVÁ 53
22. SLOVENSKÁ ŠTATISTICKÁ A DEMOGRAFICKÁ KONFERENCIA
22nd SLOVAK STATISTICS AND DEMOGRAPHIC CONFERENCE
Informácia/Information

Informácia/Information 56
NOVÝ PREDSEDA ŠTATISTICKÉHO ÚRADU SLOVENSKEJ REPUBLIKY
NEW PRESIDENT OF THE STATISTICAL OFFICE OF THE SLOVAK REPUBLIC

III. PRIPRAVUJEME/COMING SOON 57

Branislav ŠPROCHA
INFOSTAT – Výskumné demografické centrum
Centrum spoločenských a psychologických vied SAV, v.v.i.

**VZDELANIE, EKONOMICKÁ AKTIVITA A ZDRAVOTNÝ STAV OSÔB
V MARGINALIZOVANÝCH RÓMSKÝCH KOMUNITÁCH NA SLOVENSKU PODĽA
VÝSLEDKOV ZISŤOVANIA EU-SILC MRK 2020¹**

**EDUCATION, ECONOMIC ACTIVITY AND HEALTH STATUS OF PERSONS IN
MARGINALIZED ROMA COMMUNITIES IN SLOVAKIA ACCORDING TO THE
RESULTS OF THE EU-SILC MRC 2020 SURVEY**

ABSTRAKT

Cieľom príspevku je prostredníctvom výsledkov zo zisťovania EU-SILC MRK 2020 analyzovať vzdelanostnú štruktúru, ekonomické postavenie a zdravotný stav osôb žijúcich v marginalizovaných rómskych komunitách na Slovensku. Získané závery potvrdzujú skoré ukončovanie vzdelanostnej dráhy, veľmi nízku úroveň dosiahnutého vzdelania, najčastejšie základného s nezanedbateľným podielom osôb, ktoré formálne vzdelanie dosiahli v špeciálnych školách. So skorým ukončením vzdelávania je spojený skorý prechod na trh práce. Práve nízka kvalita ľudského kapitálu je jedným z aspektov vysokej nezamestnanosti, a to najmä u mladých osôb. U žien identifikujeme veľmi časté prípady pozície osoby v domácnosti, na materskej alebo rodičovskej dovolenke. S rastúcim vekom sa pomerne dynamicky zvyšuje aj zastúpenie invalidných osôb. To súvisí s kvalitou zdravotného stavu, ktorá sa vo všeobecnosti s vekom rýchlo znižuje. Týka sa to najmä prítomnosti chronických a dlhotrvajúcich ochorení, ako aj obmedzení bežných denných aktivít. Platí pritom, že o niečo dynamickejšie sa kvalita zdravia zhoršuje u žien.

ABSTRACT

The aim of the article is to analyze the educational structure, economic and health status of persons living in marginalized Roma communities in Slovakia through the results of the EU-SILC MRK 2020 survey. The obtained conclusions confirm the early termination of educational paths, a very low level of education attained most often basic with a significant share of people who achieved formal education in special schools. Early completion of education is also associated with an early transition to the labor market. The low quality of human capital is one of the aspects of high unemployment, especially among young people. In case of women, we identify very frequent cases of the position of a person in the household, on maternity or parental leave. With increasing age, the representation of disabled people also increases relatively dynamically. This is related to the quality of health, which generally declines rapidly with age. This mainly concerns the presence of chronic and protracted illness, as well as the limitation of normal daily activities. At the same time, it is true that the quality of health deteriorates somewhat more dynamically among women.

KLÚČOVÉ SLOVÁ

marginalizované rómske komunity, EU-SILC, vzdelanie, ekonomická aktivita, zdravotný stav, Slovensko

¹ Príspevok je výsledkom projektu APVV–23–0062 *Depopulácia a destabilizácia? Prognózy a simulácie demografického vývoja SR do konca 21. storočia a modelovanie jeho vybraných dopadov.*

KEY WORDS

marginalized Roma communities, EU-SILC, education, economic activity, health status, Slovakia

1. ÚVOD

V rokoch 2018 a 2020 sa Slovensku v dvoch vlnách realizovalo špecializované zisťovanie EU-SILC MRK zamerané na získanie údajov o životných podmienkach osôb žijúcich v marginalizovaných rómskych komunitách. Prebiehalo pod gesciou Štatistického úradu SR (ďalej ŠÚ SR) v spolupráci s Úradom splnomocnenca vlády pre rómske komunity (ÚSVRK). Pomocou Atlasu rómskych komunít 2019 bolo toto zisťovanie koncipované tak, aby údaje boli reprezentatívne práve pre osoby žijúce v tomto špecifickom prostredí. Údaje sú reprezentatívne za 4 typy priestorového osídlenia populáciou MRK: koncentrácia vo vnútri obce, koncentrácia na okraji obce, segregované osady a žijúci v rozptyle s majoritnou populáciou Slovenska. Zároveň sú údaje reprezentatívne z geografického aspektu za Banskobystrický, Prešovský a Košický kraj, pričom ostatné kraje predstavujú zvyšok Slovenska [5]. Práve možnosť analýzy výsledkov reprezentatívneho zisťovania za osoby žijúce v MRK na Slovensku predstavuje dôležitý a vo svojej podstate zatiaľ v našich dátových podmienkach aj jedinečný informačný vstup na empirické uchopenie niektorých štruktúrálnych vlastností danej populácie.

Cieľom predloženého príspevku je práve snaha o odhalenie charakteru vzdelanostnej a ekonomickej štruktúry osôb z MRK v kombinácii s ich vekom a pohlavím. Práve tieto dva populačné štruktúrne znaky sú vnímané ako jeden z najväčších problémov daných komunít z hľadiska socio-ekonomickej integrácie do nerómskej populácie, ako aj jeden z kľúčových dôvodov ich viacnásobného znevýhodnenia [8, 9]. Navyše marginalizácia z pohľadu dosahovania kvalitného vzdelania a možností aktívnej participácie na trhu práce sú faktory, ktoré sú úzko vzájomne podmieňujúce. S tým sa však často spája ďalší často reflektovaný atribút, akým sú nepriaznivé životné podmienky odzrkadľujúce sa na zlom zdravotnom stave osôb z MRK [3, 7, 10, 11, 12]. Keďže zisťovanie EU-SILC MRK 2020 rovnako ako „klasické“ zisťovania EU-SILC obsahovalo aj modul otázok zameraných na hodnotenie kvality zdravia, snažili sme sa v našom príspevku analyzovať aj výsledky týkajúce sa tejto problematiky.

2. CHARAKTERISTIKA ZDROJA ÚDAJOV

Základnou jednotkou zisťovania EU-SILC MRK 2020 bola hospodáriaca domácnosť, ktorú tvorili spoločne žijúce a hospodáriace osoby v jednom byte. Za znak spoločného hospodárenia sa považovala spoločná úhrada základných výdavkov domácnosti (strava, náklady za bývanie, elektrina, plyn a pod.). V rámci jednej bytovej domácnosti (spoločne bývajúce osoby) bolo možné identifikovať jednu, ale aj viac hospodáriacich domácností. Keďže zisťovanie prihliadalo aj na špecifiká rodinného správania v tomto prostredí, do domácností rodičov sa zaradili aj také nevydaté a neženaté deti, ktoré už mali svojho partnera/partnerku, ale z tohto zväzku (alebo iného predchádzajúceho) sa im ešte nenarodili žiadne deti.

Základom zisťovania boli 3 druhy dotazníkov a v roku 2020 aj doplnkový dotazník FRA. V dotazníku A označenom ako *Zloženie domácností* sa zisťovali údaje o zložení domácností, ktoré opytovateľ získal od jednotlivých osôb alebo od niektorého dospelého člena hospodáriacej domácnosti za všetkých jej členov bez ohľadu na vek.

V dotazníku B *Údaje za hospodáriacu domácnosť* opytovateľ zisťoval údaje od osoby, ktorá vedela najlepšie poskytnúť informácie o danej domácnosti. Išlo najmä o údaje týkajúce sa infraštruktúry, finančnej situácie domácnosti, dostupnosti k službám a údaje o sociálnom vylúčení. Súčasťou bol v roku 2020 aj model týkajúci sa vplyvu pandémie ochorenia COVID-19. V dotazníku C *Údaje za osoby* zisťoval vybrané údaje priamo od jednotlivých členov hospodárijacej domácnosti, ktorí k septembru 2020 mali 16 rokov a viac. Špeciálny dotazník FRA sa týkal povedomia o právach, skúsenosti s diskrimináciou, násilím, obťažovaním, políciou a dôverou v inštitúcie. Otázky vyplňal vždy jeden vybraný člen domácnosti vo veku 16 a viac rokov. Odpovede sa týkali len samotného respondenta.

Na ciele nášho príspevku bol dôležitý predovšetkým dotazník A a C. Údaje z oboch sú voľne dostupné na stránkach ÚSVRK v databázach označených ako R a P. Konkrétne z prvého menovaného pracujeme s údajmi týkajúcimi sa veku, pohlavia a najvyššieho dosiahnutého vzdelania. Zisťovanie vzdelania sa opieralo o pomerne podrobnú klasifikáciu, keď najvyššie dosiahnuté vzdelanie predstavovalo najvyšší dosiahnutý stupeň vo vzdelávacom systéme spojený so získaním príslušného vysvedčenia, certifikátu alebo diplomu. Tie osoby, ktoré v čase zisťovania boli aktívne vo vzdelávacom procese (vzdelávali sa, navštevovali školu), mali uviesť svoje posledné dosiahnuté vzdelanie (a nie práve „študovaný“ stupeň). V zisťovaní EU-SILC MRK 2020 sa pracovalo s nasledujúcou klasifikáciou: 1. bez školského vzdelania; 2. základné vzdelanie – a) základné 1. stupeň, b) základné 1. stupeň špeciálne vzdelávanie, c) základné 2. stupeň, d) základné 2. stupeň špeciálne vzdelávanie; 3. stredné odborné (učňovské) vzdelanie bez maturity – a) bez výučného listu (zaškolenie, zaučenie), b) s výučným listom, c) s výučným listom MP (mentálne postihnutý), d) s vysvedčením o záverečnej skúške; 4. úplné stredné vzdelanie s maturitou – a) odborné (učňovské) s výučným listom, b) všeobecné, c) odborné, d) z nadstavbového štúdia; 5. vyššie odborné a pomaturitné vzdelanie – a) vyššie odborné (pomaturitné), b) vyššie odborné (neuniverzitné, absolventské); 6. vysokoškolské vzdelanie – a) vysokoškolské – 1. stupeň (Bc.), b) vysokoškolské – 2. stupeň (Mgr., Ing., MUDr., JUDr. a iné), c) vysokoškolské – 3. stupeň (PhD. a iné).

Vzhľadom na veľmi nízke zastúpenie osôb s úplným stredoškolským a vyšším vzdelaním sme tieto skupiny zlúčili do jednej umelej označenej ako vyššie vzdelanie.

Celkovú vzdelanostnú úroveň osôb z prostredia MRK a najmä jej kvalitatívnu stránku sme sa snažili vyjadriť aj prostredníctvom niektorých syntetických ukazovateľov [bližšie napr. 6]. Konkrétne sme sa opierali o vzdelanostný potenciál a priemernú dĺžku vzdelávania. Vzdelanostný potenciál je zostavený tak, aby vyjadroval priemernú úroveň vzdelanosti analyzovanej populácie. Základnom jeho konštrukcie je, že jednotlivým stupňom dosiahnutého vzdelania sa priradia koeficienty od 0 v prípade osôb bez vzdelania po 17 pre osoby s tretím stupňom vysokoškolského vzdelania. Zachovaná tak je logika, že čím vyššie vzdelanie, tým vyšší koeficient sa mu priradí. Vo svojej podstate indikátor prezentuje priemernú kvalitu dosiahnutého vzdelania. Vo všeobecnosti pritom platí, že čím je číslo vyššie, tým je vyššia priemerná kvalita vzdelanostnej štruktúry. Vzdelanostný potenciál sme zostavovali zvlášť pre mužov a ženy a jednotlivé 5-ročné vekové skupiny. Ďalším zo spôsobov, ako prostredníctvom syntetického ukazovateľa vyjadriť „kvalitu“ dosahovaného vzdelania je použiť dĺžku formálneho vzdelávacieho procesu. Tá predstavuje počet rokov, ktoré je potrebné štandardne stráviť vo formálnom vzdelávaní na dosiahnutie určitého

stupňa vzdelania. Z údajov o štruktúre obyvateľstva podľa dosiahnutého vzdelania je potom možné odvodiť priemernú dĺžku vzdelania. Pri konštrukcii tohto indikátora je každý stupeň formálneho štúdia kvalifikovaný prislúchajúcim počtom rokov jeho štandardnej dĺžky [6].

Z dotazníka C (databáza P) využívame otázky týkajúce sa veku, pohlavia a súčasného ekonomického postavenia² osoby vo veku 16 a viac rokov. Išlo pritom o samodefinovaný, v čase zisťovania aktuálny status ekonomickej aktivity identifikovaný na základe najviac stráveného času. Konkrétne v zisťovaní EU-SILC MRK 2020 sa využili nasledujúce kategórie ekonomického statusu: 1.) zamestnanec pracujúci na plný pracovný čas, 2.) zamestnanec pracujúci na kratší pracovný čas, 3.) podnikateľ, samostatne zárobkovo činná osoba na plný pracovný čas (vrátane vypomáhajúcich členov domácnosti v rodinnom podniku), 4.) podnikateľ, samostatne zárobkovo činná osoba na kratší pracovný čas (vrátane vypomáhajúcich členov domácnosti v rodinnom podniku), 5.) nezamestnaný, 6.) žiak, študent, ďalšie vzdelávanie, neplatené doškoľovanie, 7.) starobný dôchodca, osoba v predčasnom dôchodku, 8.) invalidná osoba alebo osoba nespôsobilá pracovať, 9.) osoba v domácnosti, 10.) iná neaktívna osoba.

Zaujímavým prvkom zisťovania EU-SILC MRK 2020 bola tiež otázka na vek nástupu respondenta do prvého zamestnania. Prostredníctvom nej sme sa snažili identifikovať načasovanie prechodov na pracovný trh. Za prvé zamestnanie sa považovalo také zamestnanie, ktoré trvalo najmenej 6 po sebe idúcich mesiacov. Príležitostné práce, ako aj práce počas štúdia sa za prvé zamestnanie nepočítali. Určitou výnimkou bola situácia, keď sa zamestnanie začalo ešte počas štúdia a kontinuálne pokračovalo aj po jeho skončení s celkovou dĺžkou spomínaných minimálne 6 mesiacov.

Veľmi dôležitou súčasťou dotazníka C a databázy P boli tiež otázky zamerané na niektoré dimenzie kvality zdravotného stavu respondentov. V prvej³ je snaha získať obraz o všeobecnom (nie aktuálnom, platnom v čase zisťovania) charaktere zdravotného stavu. Ide o sebareflexiu kvality vlastného zdravotného stavu na 5-stupňovej škále: 1.) veľmi dobré, 2.) dobré, 3.) ani dobré ani zlé, 4.) zlé, 5.) veľmi zlé.

Na účely našej analýzy boli dané odpovede integrované do troch hlavných skupín: 1.) veľmi dobré a dobré, 2.) ani dobré ani zlé, 3.) zlé a veľmi zlé.

V druhej nami využívanej otázke⁴ sa zisťovala prítomnosť chronického ochorenia. Ako dlhodobé ochorenie alebo zdravotný problém sa rozumelo také ochorenie alebo problém, ktoré pretrvávali, alebo sa očakávalo, že budú pretrvávať viac ako 6 mesiacov alebo dlhšie. Zdravotný problém pritom nemusel byť nevyhnutne diagnostikovaný lekárom. Dlhodobý zdravotný problém mohol vzniknúť následkom poranenia, úrazov, vrodených chýb a chorôb. Patrili sem aj opakované problémy, ktoré sú sezónne alebo nesúvislé, aj keď sa prejavujú počas menej ako 6 mesiacov (napr. alergie). Dočasné (prechodné) zdravotné problémy sa nebrali do úvahy.

² Osoby vo veku 16 a viac rokov odpovedali na nasledujúcu otázku: „Aké je Vaše súčasné ekonomické postavenie?“

³ Respondenti odpovedali na otázku: „Ako by ste celkovo (vo všeobecnosti) zhodnotili Vaše zdravie?“

⁴ Respondenti prítomnosť dlhodobého ochorenia alebo zdravotného problému reflektovali otázkou: „Máte nejaké dlhodobé ochorenie alebo dlhodobý zdravotný problém?“

Z hľadiska miery obmedzovania bežných činností⁵ respondent sám zhodnotil, či je obmedzovaný nejakými pretrvávajúcimi fyzickými alebo mentálnymi zdravotnými problémami, chorobou alebo invaliditou oproti všeobecnému štandardu populácie. Do úvahy boli brané len obmedzenia zo zdravotných dôvodov a nie obmedzenia podmienené finančnými, kultúrnymi alebo inak so zdravotným stavom nesúvisiacimi faktormi. Zahrnuté boli len také prípady, keď trvanie obmedzenia činnosti začalo minimálne pred 6 mesiacmi a pretrvávalo počas zisťovania. Z hľadiska miery obmedzovania sa mal respondent následne vyjadriť, či je: 1.) veľmi obmedzovaný, 2.) obmedzovaný ale nie veľmi, 3.) vôbec neobmedzovaný.

Do prvej skupiny sa mali respondenti zaradiť v prípade, ak aktivity, ktoré ľudia bežne robia, vykonávajú len s veľkými ťažkosťami alebo ich nedokážu vykonávať takmer vôbec a je potrebná asistencia inej osoby. V druhej skupine sa nachádzali respondenti, ktorí bežné denné aktivity dokázali vykonávať s určitými ťažkosťami. Pri ich vykonávaní nepotrebovali pomoc od inej osoby, a ak áno, nebolo to na dennej báze. Do poslednej tretej skupiny sa zaradili respondenti, ktorí bežné denné aktivity dokážu vykonávať bez akýchkoľvek ťažkostí alebo obmedzenie pri ich vykonávaní trvá menej ako 6 mesiacov, teda toto obmedzenie zatiaľ nebolo dlhodobého charakteru.

3. VZDELANOSTNÁ ŠTRUKTÚRA

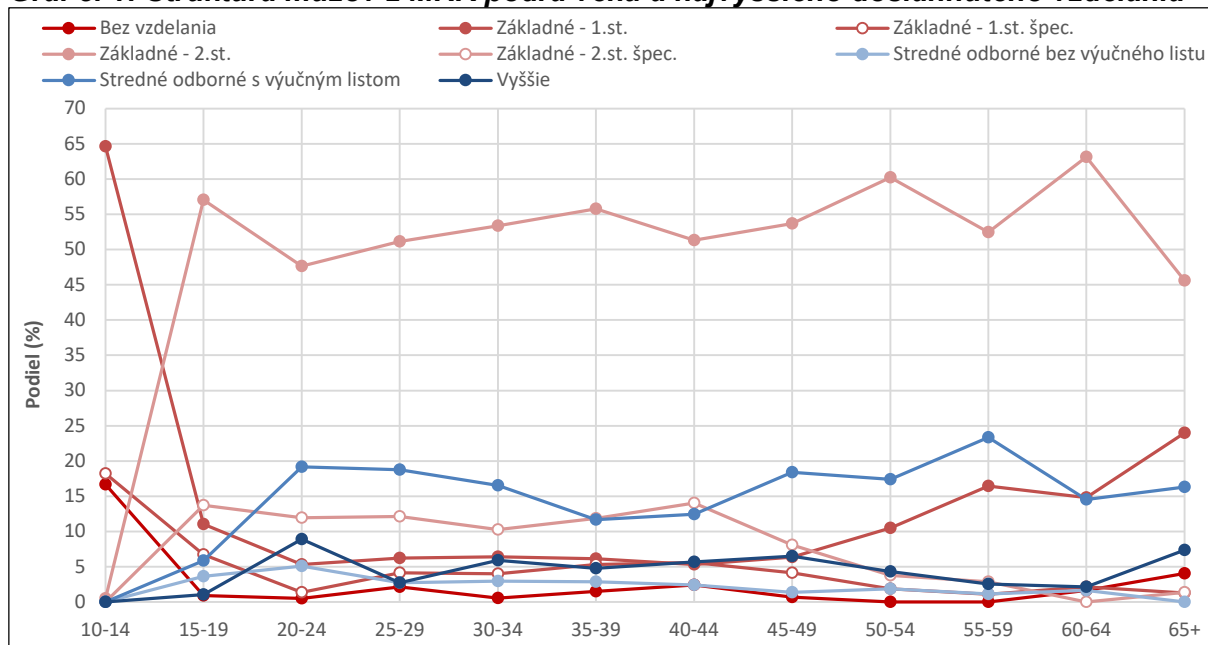
Úroveň dosiahnutého vzdelania Rómov a obzvlášť osôb z MRK je na Slovensku dlhodobo identifikovaná ako veľmi nízka [14]. Kľúčovým problémom je predovšetkým skutočnosť, že väčšine mužov a žien z prostredia MRK končia ich vzdelanostné dráhy veľmi skoro a najčastejším vzdelanostným modelom u dospelých je ukončenie základnej školy. Len veľmi malá časť získava stredoškolské vzdelanie, a aj to maximálne bez maturity [14].

Výsledky zisťovania EU-SILC MRK 2020 tento model potvrdili. Vzhľadom na veľmi nízke počty osôb s maturitou a absolventov terciárneho stupňa vzdelania sme tieto vzdelanostné skupiny zlúčili a vytvorili sme umelú kategóriu vyššie vzdelanie. Aj napriek tomu tento vzdelanostný stupeň dosahovalo u oboch pohlaví menej ako 10 % osôb. Najbližšie sa k tejto hranici priblížili muži i ženy vo veku 20 – 24 rokov. S výnimkou najmladšej analyzovanej vekovej skupiny (10 – 14 rokov) mali u mužov i žien jednoznačnú prevahu osoby, ktoré v zisťovaní deklarovali, že ukončili druhý stupeň základného vzdelania. U mužov sa ich podiel pohyboval v rozmedzí približne 45 – 60 % (graf č. 1) a u žien to bolo dokonca 50 – 70 % (graf č. 2). S výnimkou najstaršej seniorskej vekovej skupiny sa dá pritom povedať, že podiel osôb s týmto stupňom vzdelania mal mierne rastúci trend. V starších vekových skupinách sa zvyšovalo aj zastúpenie mužov a žien, ktorí dosiahli maximálne prvý stupeň základnej školy. Vo veku 65 a viac rokov napríklad u mužov tvorili takmer štvrtinu a u žien dokonca 27 %.⁶

⁵ Respondent odpovedal na otázku: „Do akej miery ste najmenej posledných 6 mesiacov obmedzovaný kvôli zdravotnému problému v činnostiach, ktoré ľudia bežne vykonávajú?“

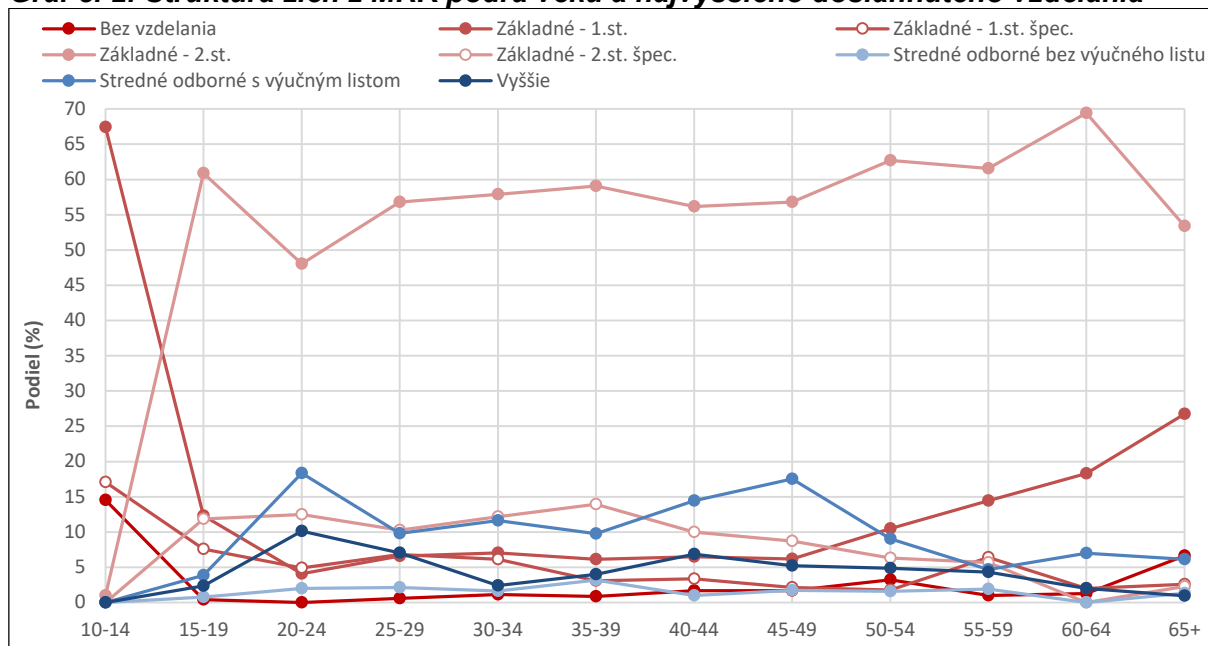
⁶ Pre úplnosť ešte doplníme, že na Slovensku v produktívnom veku tvorili podľa výsledkov SODB 2021 muži so základným vzdelaním len niečo viac ako 11 %. U žien bolo zastúpenie tejto vzdelanostnej skupiny len o niečo vyššie (12 %). Ďalších 42 % (muži) resp. takmer 30 % (ženy) predstavovali osoby so stredoškolským vzdelaním bez maturity. Vyššie vzdelanie vo veku 15 – 64 rokov dosahovalo približne 46 % mužov a 58 % žien.

Graf č. 1: Štruktúra mužov z MRK podľa veku a najvyššieho dosiahnutého vzdelania



Zdroj: EU-SILC MRK 2020; výpočty autora

Graf č. 2: Štruktúra žien z MRK podľa veku a najvyššieho dosiahnutého vzdelania



Zdroj: EU-SILC MRK 2020; výpočty autora

Ako ukazuje graf č. 1 a 2, podľa výsledkov zisťovania nezanedbateľnou súčasťou vzdelanostnej dráhy osôb z MRK na Slovensku je, že základné vzdelanie absolvujú v prostredí špeciálnych škôl. U oboch pohlaví sa podiel týchto osôb vo veku do 50 rokov pohyboval v rozmedzí 15 – 20 %. Zaujímavosťou pritom je, že v starších vekových skupinách ich zastúpenie klesalo, a to najmä v prospech vzdelávania v školách bez špecializovanej formy vzdelávania.

Ak muž alebo žena z MRK dosiahli vyššie ako základné vzdelanie, najčastejšie išlo o stredné odborné bez výučného listu. Podiel mužov s týmto typom vzdelania sa vo veku nad 20 rokov pohyboval v rozmedzí 10 – 25 %. Platilo pritom, že vyššie

zastúpenie (tesne pod 20 %) identifikujeme skôr v mladšom veku (20 – 29 rokov) a najnižšie vo veku 35 – 44 rokov. Následne v starších produktívnych vekových skupinách (60 – 64 rokov) ich podiel opätovne rastie až takmer k jednej štvrtine.

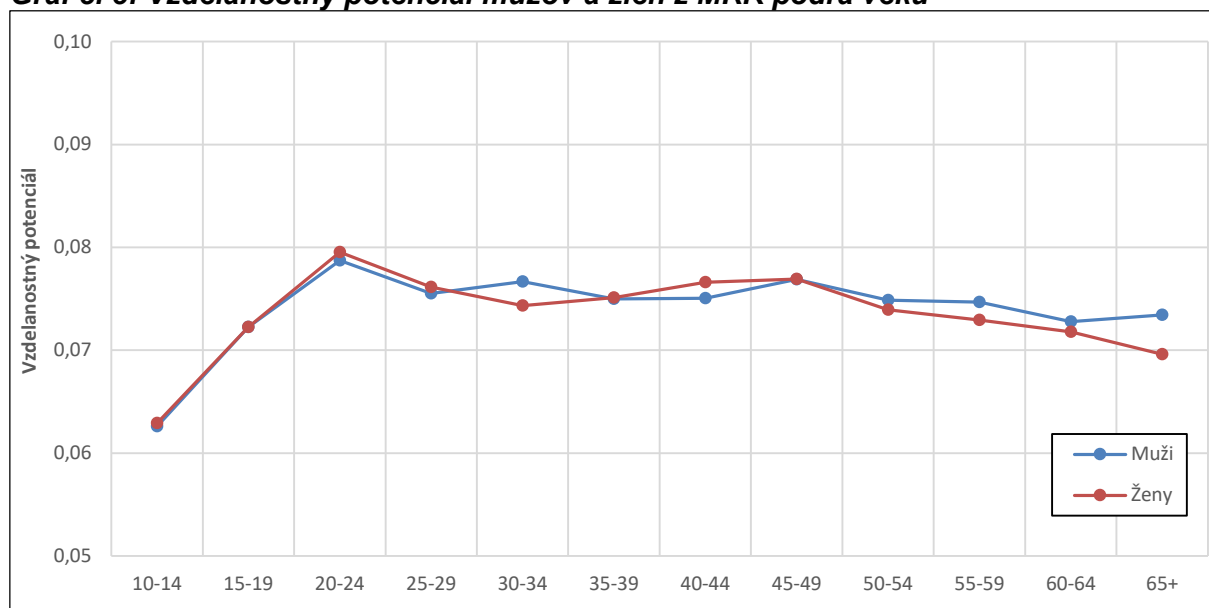
U žien je podiel osôb so stredným vzdelaním s výučným listom celkovo o niečo nižší, keďže tvoria medzi 20. a 55. rokom života 10 – 20 %. V mladších a starších vekových skupinách ich zastúpenie klesá pod túto hranicu a pohybuje sa okolo úrovne 5 %. K hornej hranici spomínaného intervalu sa však ženy z MRK dostávajú len vo veku 20 – 24 a 45 – 49 rokov.

Pre úplnosť ešte dodávame, že podľa výsledkov zisťovania EU-SILC MRK 2020 podiel osôb so stredným vzdelaním bez výučného listu je u oboch pohlaví veľmi nízky a dosahuje maximálne 5 % u mužov (20 – 24 rokov) a 3 % u žien (35 – 39 rokov). Obdobne s výnimkou veku do 15 rokov je marginálne aj zastúpenie osôb bez vzdelania. V seniorskom veku sa ich podiel u mužov pohybuje v priemere na úrovni 4 % a u žien tvorí približne 7 %.

Veľmi nepriaznivú vzdelanostnú štruktúru osôb z MRK potvrdzujú aj niektoré syntetické ukazovatele snažiace sa empiricky vyjadriť jej celkovú kvalitatívnu stránku [bližšie 6]. Vzdelanostný potenciál je zostavený tak, aby vyjadroval priemernú úroveň vzdelanosti analyzovanej populácie. Základom jeho výpočtu je, že jednotlivým stupňom dosiahnutého vzdelania sa priradia koeficienty od 0 u osôb bez vzdelania po 17 pre osoby s tretím stupňom vysokoškolského vzdelania. Zachovaná tak je logika, že čím vyššie vzdelanie, tým vyšší koeficient sa mu priradí. Vo svojej podstate indikátor prezentuje priemernú kvalitu dosiahnutého vzdelania, pričom platí, že čím je číslo vyššie, tým je vyššia priemerná kvalita vzdelanostnej štruktúry. Výsledné hodnoty vzdelanostného potenciálu podľa vekových skupín osôb z MRK pre mužov a ženy prezentuje graf č. 3. Ako je z neho zrejmé, od najmladšieho veku najprv dochádza logicky k určitému nárastu s maximom vo veku 20 – 24 rokov. Smerom k staršiemu veku sa vzdelanostný potenciál postupne mierne znižuje, čo signalizuje znižujúcu sa kvalitu dosiahnutého vzdelania starších osôb. Diferencie medzi pohlaviami sa ukazujú byť relatívne malé s výnimkou najstarších vekových kohort, kde predsa len identifikujeme o niečo „kvalitnejšiu“ vzdelanostnú štruktúru na strane mužov z MRK.

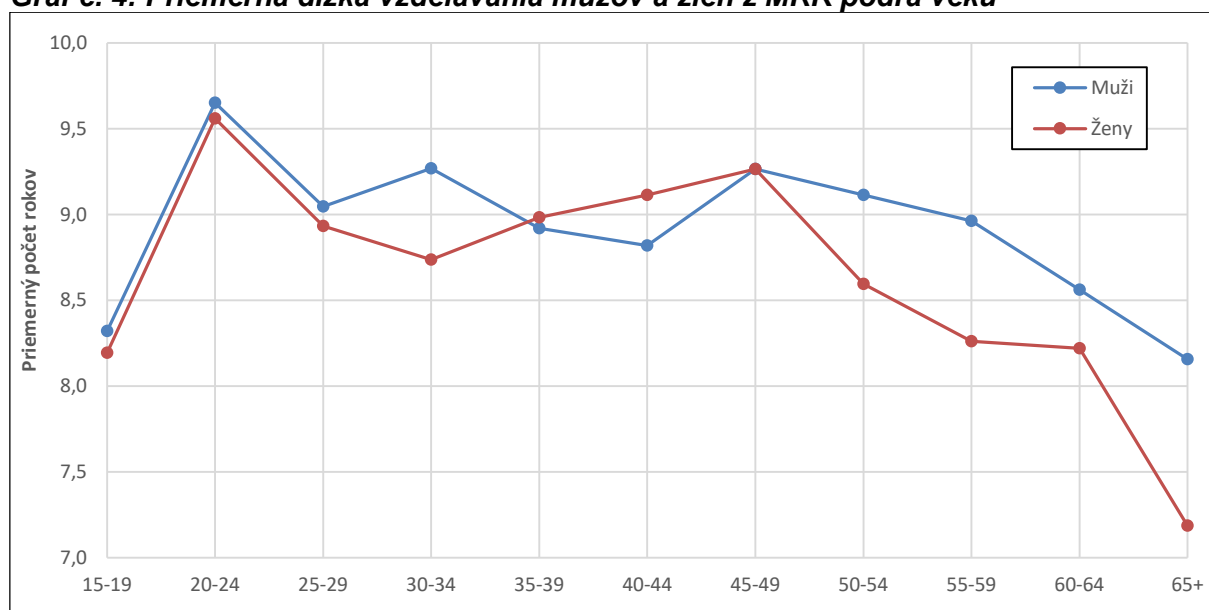
Ďalším zo spôsobov, ako prostredníctvom syntetického ukazovateľa vyjadriť „kvalitu“ dosahovaného vzdelania je použiť dĺžku vzdelávania. Tá predstavuje počet rokov, ktoré je potrebné štandardne stráviť vo formálnom vzdelávaní na dosiahnutie určitého stupňa vzdelania. Z údajov o štruktúre obyvateľstva podľa dosiahnutého vzdelania je potom možné odvodiť priemernú dĺžku vzdelávania. Pri konštrukcii tohto indikátora je každý stupeň formálneho štúdia kvalifikovaný prislúchajúcim počtom rokov jeho štandardnej dĺžky [6]. Údaj o priemernej dĺžke vzdelávania mužov a žien z MRK podľa výsledkov zisťovania EU-SILC MRK 2020 prezentuje graf č. 4. Aj v tomto prípade platí, že „najkvalitnejšie“ vzdelanie v prostredí MRK dosahovali mladšie osoby, o niečo skôr muži, no v staršom veku a najmä seniorskom sa priemerná dĺžka vzdelávania signifikantne skracovala.

Graf č. 3: Vzdelanostný potenciál mužov a žien z MRK podľa veku



Zdroj: EU-SILC MRK 2020; výpočty autora

Graf č. 4: Priemerná dĺžka vzdelávania mužov a žien z MRK podľa veku



Zdroj: EU-SILC MRK 2020; výpočty autora

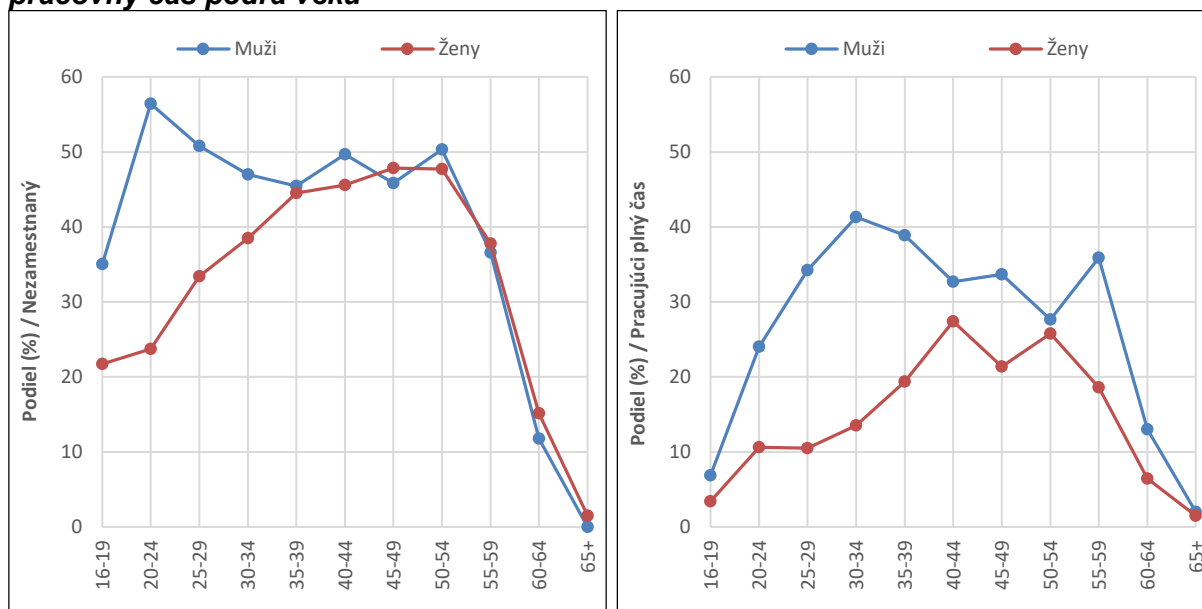
4. EKONOMICKÉ POSTAVENIE

Pozícia osôb rómskeho pôvodu na pracovnom trhu je vo všeobecnosti vnímaná ako problematická [4, 8, 9]. Práve Rómovia tvoria významnú časť dlhodobo nezamestnaných osôb na Slovensku. Na vysoký podiel nezamestnaných osôb v prostredí MRK poukazujú aj výsledky zisťovania EU-SILC MRK 2020. U mužov v podstate s výnimkou najmladšieho veku do 20 rokov a následne vo veku nad 55 rokov sa podiel nezamestnaných pohyboval okolo hranice 50 %. Jednoznačne najvyšší pritom bol u osôb tesne po skončení školy vo veku 20 – 24 rokov, keď dosahoval približne 56 %. Situácia u žien je predovšetkým v reprodukčnom veku odlišná. Súvisí to so skorším začiatkom materskej dráhy (pozri ďalej) a teda s častejšou prítomnosťou žien v domácnosti alebo žien patriacich do kategórie inak neaktívnych. Podľa metodických pokynov na vyplňovanie dotazníka sem opytovateľa

mali zaradiť napríklad ženy na rodičovskej dovolenke. Na konci a po skončení reprodukčného obdobia sa však podiel nezamestnaných žien výraznejšie neodlišoval od zastúpenia nezamestnaných mužov (graf č. 5). Vzhľadom na rastúci podiel osôb v predčasnom dôchodku, starobnom dôchodku, ako aj invalidných osôb podiel nezamestnaných vo veku 55 a viac rokov pomerne rýchlo klesá. V seniorskom veku je ich zastúpenie minimálne (graf č. 5).

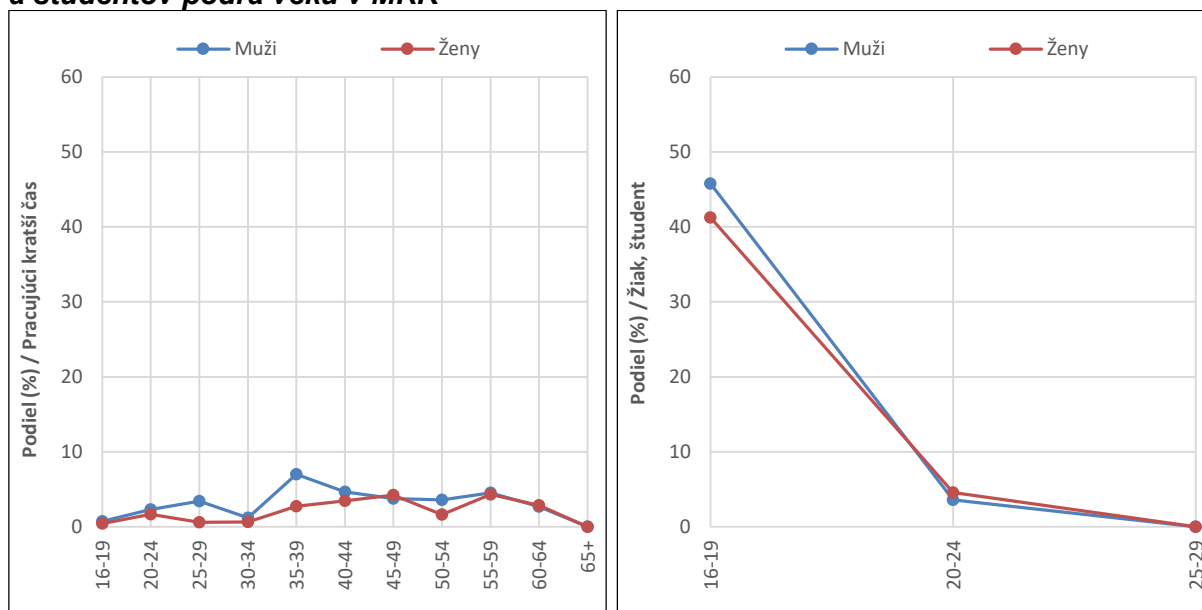
Zisťovanie EU-SILC MRK 2020 pracovalo z hľadiska aktuálnej ekonomickej aktivity s 2 základnými skupinami pracujúcich (zamestnanec alebo podnikateľ, samostatne zárobkovo činná osoba vrátane vypomáhajúcich členov domácnosti v rodinnom podniku), rozdelených do dvoch podskupín podľa dĺžky pracovného času (plný, kratší pracovný čas). Pri našej analýze sme ich zlúčili do 2 skupín pracujúcich podľa dĺžky pracovného času. Ako je však ale z grafu č. 7 zjavné, podiel mužov a najmä žien pracujúcich na kratší pracovný čas je veľmi nízky a ani v jednej vekovej skupine neprekračuje hranicu 10 %. V prípade osôb pracujúcich na plný pracovný čas je síce situácia výrazne lepšia, ale aj napriek tomu sa potvrdzuje pomerne nízka aktívna participácia mužov a najmä žien z prostredia MRK na oficiálnom pracovnom trhu. Muži pracujúci na plný čas dosahujú najvyšší podiel vo veku 30 – 34 rokov, keď tesne prekračujú hranicu 40 % z danej vekovej skupiny. Tesne pod touto úrovňou sa nachádzajú aj v nasledujúcom veku 35 – 39 rokov. Smerom k mladším aj starším vekom je však ich zastúpenie signifikantne nižšie. Napríklad vo veku 20 – 24 rokov tvoria pracujúci na plný pracovný čas len necelú štvrtinu z celkovej populácie mužov z MRK v tomto veku. Približne tretinový podiel dosahujú vo veku 40 – 49 rokov. Po dovŕšení 60. roku života pritom pracuje bez ohľadu na dĺžku pracovného času len asi 15 % mužov.

Graf č. 5 a 6: Podiel nezamestnaných a pracujúcich mužov a žien v MRK na plný pracovný čas podľa veku



Zdroj: EU-SILC MRK 2020; výpočty autora

Graf č. 7 a 8: Podiel mužov a žien pracujúcich na kratší pracovný čas a podiel žiakov a študentov podľa veku v MRK



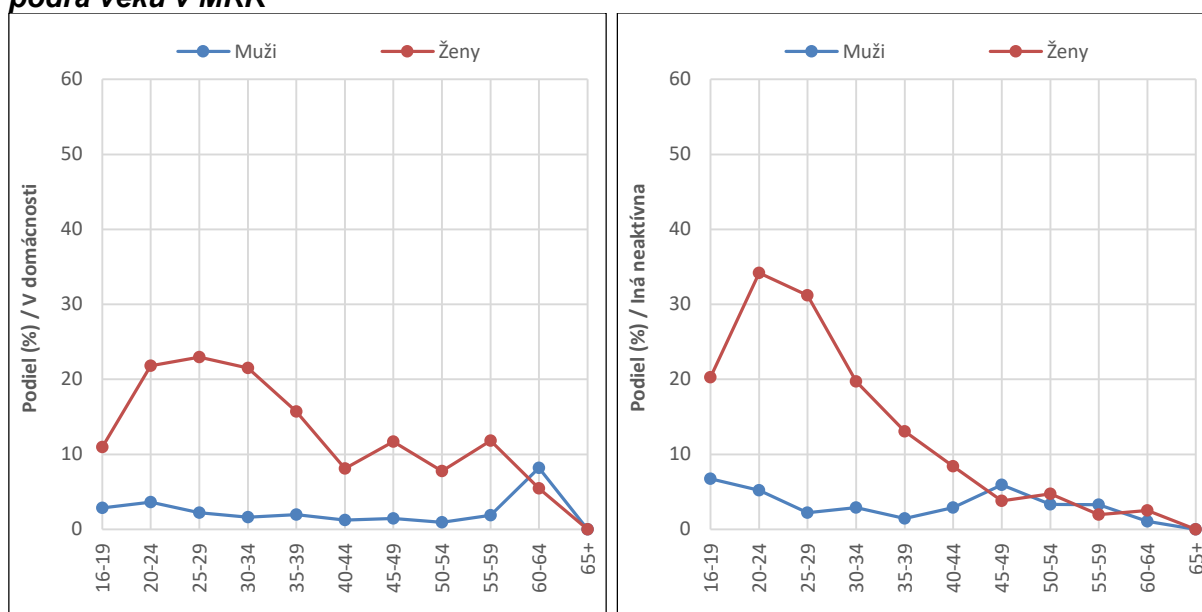
Zdroj: EU-SILC MRK 2020; výpočty autora

Podiel pracujúcich žien na plný pracovný čas je ešte nižší (graf č. 6). V mladom veku (20 – 29 rokov) je to len niečo viac ako desatina⁷, pričom najvyššie zapojenie žien do pracovného procesu pozorujeme až na konci a po skončení reprodukčného obdobia (40 – 54 rokov). V tomto vekovom spektre sa však podiel pracujúcich žien na plný úväzok pohyboval len v rozmedzí 20 – 30 % (graf č. 6). S rastúcim vekom navyše pomerne dynamicky klesá a po dovŕšení 60. roku života ide len o približne každú desiatu osobu. Výsledky zisťovania EU-SILC MRK 2020 okrem iného potvrdzujú aj veľmi skoré ukončovanie vzdelanostnej dráhy mladých mužov a žien z prostredia MRK. Nepriamo na to poukazovali už predchádzajúce analyzované údaje o vzdelanostnej štruktúre, no definitívne potvrdenie môžeme vidieť na základe podielu žiakov a študentov podľa veku a pohlavia (graf č. 8). Vo veku 16 – 19 rokov ich zastúpenie už netvorí ani polovicu z príslušnej vekovej skupiny, pričom vo veku 20 – 24 rokov to sú zhruba 4 % mužov a 5 % žien, ktoré ďalej pokračujú vo vzdelávaní. Predovšetkým doménou žien vo veku 20 – 34 rokov je častejšia starostlivosť o domácnosť. Ide pritom o osoby, ktoré nemajú platenú prácu alebo osoby, ktoré nie sú zamestnané, ale sú schopné práce, nepripravujú sa na budúce povolanie a ani si nehládajú prácu, ale sa starajú o domácnosť a osoby v domácnosti. V spomenutej vekovej skupine sa ich podiel pohyboval nad hranicou 20 % (graf č. 9). Ešte vo veku 35 – 39 rokov to bolo približne 16 %, no v staršom veku išlo približne už len o každú desiatu ženu a vo vekoch nad 60 rokov už jednoznačne predstavovali marginálny jav. Pomerne vysoké zastúpenie dosahovala u žien aj kategória iná neaktívna. Ako sme už naznačili, sem spadali napríklad osoby na rodičovskej dovolenke. To vysvetľuje prečo ich podiel dosahoval u žien najvyššie hodnoty (31 – 34 %) vo veku 20 – 29 rokov a na hranici 20 % sa pohyboval aj vo veku 16 – 19 a 30 – 34 rokov. Nielen skorý začiatok materskej dráhy, ale aj častejšia prítomnosť detí druhého a vyššieho poradia v domácnostiach prispievajú k opakovaniu rodičovskej dovolenky. Celkovo horšie

⁷ Medzi pracujúce osoby sa pritom zaraďujú aj osoby na materskej dovolenke. Nepriamo to poukazuje na skutočnosť, že drvivá väčšina žien z prostredia MRK nemá nárok na materskú dovolenku, keďže pred narodením dieťaťa nespĺňa podmienky na jej čerpanie.

úmrtnostné pomery [1, 2, 11, 13], ako aj zdravotný stav u osôb žijúcich v MRK [3, 7, 11, 13] sú pravdepodobne dôležitým faktorom postupne s vekom rastúceho podielu invalidných mužov a žien (graf č. 11). Za invalidnú osobu alebo osobu nespôsobilú pracovať sa pre potreby zisťovania EU-SILC MRK 2020 považoval taký muž alebo žena, ktorí pre dlhodobu nepriaznivý zdravotný stav mali pokles schopnosti vykonávať zárobkovú činnosť o viac ako 40 % v porovnaní so zdravou fyzickou osobou. Ako ukazuje graf č. 11, maximálne zastúpenie tieto osoby dosahovali u oboch pohlaví vo veku 55 – 59 rokov, keď v mužskej časti išlo o necelú pätinu a v ženskej časti populácie MRK dokonca o približne štvrtinu danej vekovej skupiny. Invalidita alebo neschopnosť pracovať sa však netýkala len staršieho veku, ale nezanedbateľnú časť nachádzame napríklad aj na konci reprodukčného obdobia, keď približne desatina mužov a žien podľa výsledkov EU-SILC MRK 2020 bola klasifikovaná do tejto skupiny.⁸

Graf č. 9 a 10: Podiel mužov a žien v domácnosti a so statusom inej neaktívnej osoby podľa veku v MRK



Zdroj: EU-SILC MRK 2020; výpočty autora

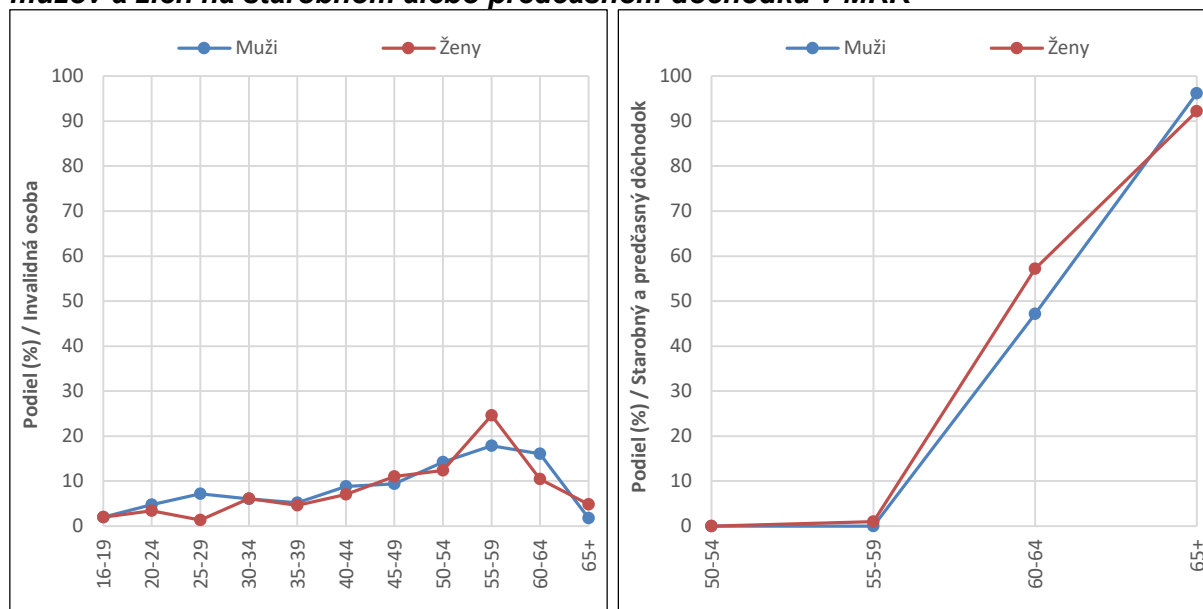
Starobný a predčasný dôchodok čerpalo podľa výsledkov zisťovania už takmer 60 % žien z MRK vo veku 60 – 64 rokov. V prípade mužov to bola menej ako polovica. Zostávajúca časť osôb je na základe vyššie uvedených štruktúr ekonomickej aktivity rozdelená u mužov na pracujúcich (16 %), nezamestnaných (12 %), invalidov a neschopných práce (16 %) a mužov v domácnosti alebo inak neaktívnych (10 %). U žien vo veku 60 – 64 rokov je vzhľadom na vyššie zastúpenie osôb v starobnom a predčasnom dôchodku (aj v spojitosti s vyšším počtom vychovaných detí) zloženie niečo viac ako 40 % osôb takého: približne 9 % v tomto veku ešte pracuje, 15 % je nezamestnaných, 10 % invalidov alebo neschopných práce a 8 % žien v tomto vekovom spektre je v domácnosti alebo patrí do skupiny iných neaktívnych. Vo veku 65 a viac rokov u mužov starobní dôchodcovia tvoria už viac ako 96 %. Zvyšok pripadal

⁸ Špecifickosť ekonomickej štruktúry osôb z MRK jednoznačne potvrdzuje aj jej porovnanie s celou populáciou Slovenska. Podľa údajov SODB 2021 v produktívnom veku tvorili na Slovensku pracujúci muži viac ako 67 % a pracujúce ženy viac ako 55 %. Ďalších približne 5 % predstavovali pracujúci dôchodcovia. Podiel osôb seba reflektujúcich ako nezamestnaných dosahoval približne 6 % u mužov a necelých 7 % u žien. Osoby v domácnosti na strane mužov tvorili 5 % a u žien takmer 12 %. Na dôchodcov vo veku 15 – 64 rokov pripadalo necelých 7 % u mužov a 8 % u žien.

na pracujúcich (2 %) a invalidné osoby (necelé 2 %). U žien zo starobného dôchodku žije viac ako 92 % osôb. Zvyšok (takmer 5 %) pripadá na invalidné alebo práceneschopné osoby a ďalšie 3 % si približne rovnako delia pracujúce a nezamestnané ženy.

Súčasťou zisťovania EU-SILC MRK 2020 je tiež otázka o veku nástupu do prvého zamestnania.⁹ Zástupcovia oboch pohlaví najčastejšie uvádzali veko 18 rokov (graf č. 13). Platí však, že výrazne častejšie túto hranicu deklarovali muži, keďže išlo o približne tretinu odpovedí. U žien tento vek tvoril necelých 22 % všetkých odpovedí. Nad hranicou 10 % sa u mužov zistili aj ďalšie dva veku. Relatívne častejšie boli aj v prípade žien, no opätovne nedosahovali takú frekvenciu ako v mužskej časti populácie MRK. Ako je však z grafu č. 13 zrejmé, u žien bol naopak o niečo častejší vyšší vek, v ktorom respondentky deklarovali svoje prvé zamestnanie. Išlo napríklad o vek 25, 30 alebo 35 rokov. Rovnako vyššie zastúpenie ako u mužov mal aj vek 40 a viac rokov. Zaujímavý je pritom efekt vyššej frekvencie rokov končiacich sa 0, 5, čo vzbudzuje podozrenie, že respondentky uvádzali skôr svoj odhad, teda približný ako presný vek začiatku prvého pravidelného zamestnania. Súčasne však častejšie uvádzanie vyššieho veku môže odzrkadľovať realitu, keďže viaceré zo žien žijúcich v prostredí MRK sa do pracovného procesu dostávajú až po skončení reprodukčnej dráhy, resp. reťazových pôrodov, a s tým súvisiacej rodičovskej dovolenky a starostlivosti o deti a domácnosť. Potvrďuje to napokon aj priemerný vek mužov a žien pri prvom pravidelnom zamestnaní. Podľa reflektovaných údajov v zisťovaní EU-SILC MRK 2020 sa u mužov pohyboval na úrovni niečo viac ako 21 rokov, kým u žien dosahoval až 25 rokov.

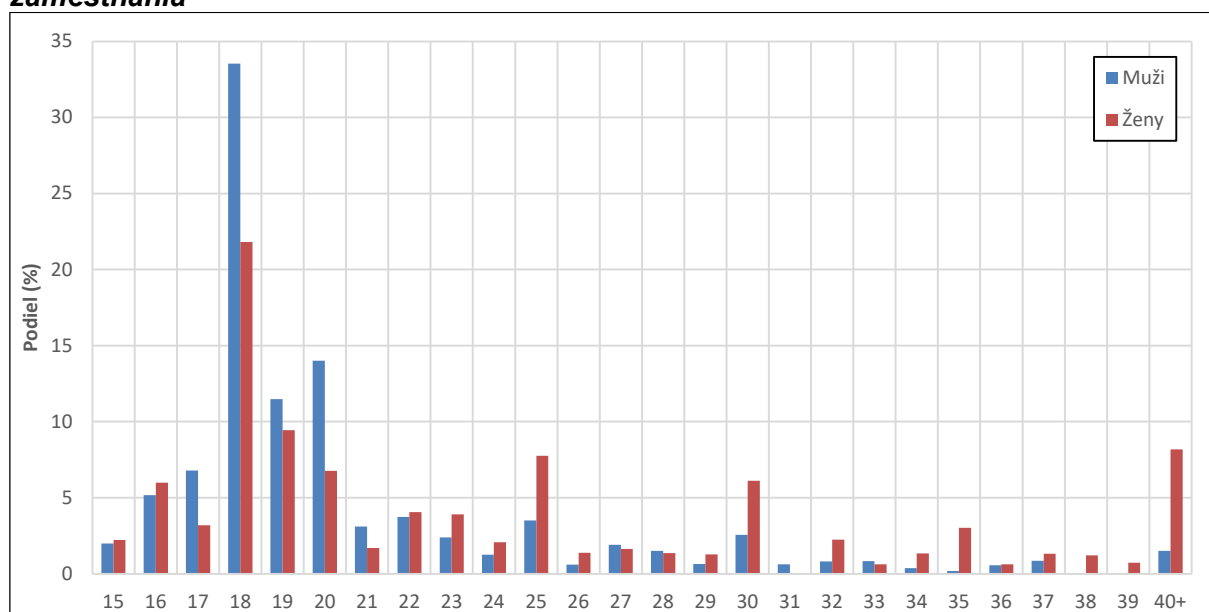
Graf č. 11 a 12: Podiel invalidných alebo inak práceneschopných mužov a žien a podiel mužov a žien na starobnom alebo predčasný dôchodku v MRK



Zdroj: EU-SILC MRK 2020; výpočty autora

⁹ Koľko ste mali rokov, keď ste začali vykonávať svoje prvé pravidelné zamestnanie alebo podnikanie?

Graf č. 13: Štruktúra odpovedí mužov a žien z MRK na vek začiatku prvého pravidelného zamestnania



Zdroj: EU-SILC MRK 2020; výpočty autora

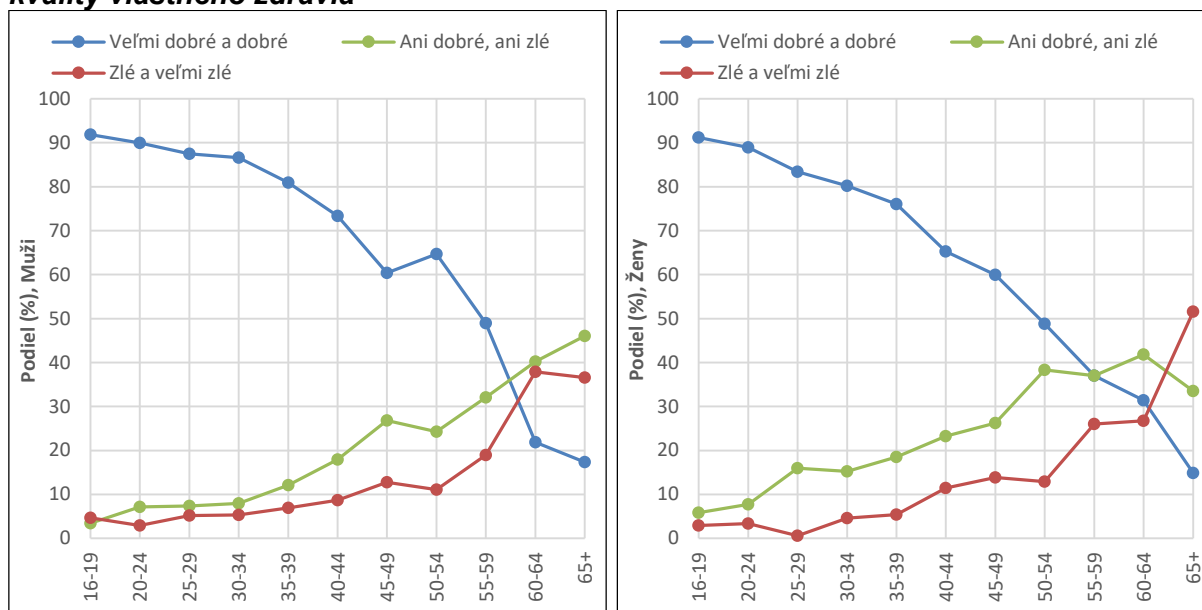
5. ZDRAVOTNÝ STAV

Dôležitou súčasťou zisťovania EU-SILC MRK 2020 bol modul zameriavajúci sa na zdravotný stav osôb z prostredia MRK. Ten sa opiera o sériu troch otázok zisťujúcich jednak subjektívne hodnotenie kvality vlastného zdravotného stavu, prítomnosť chronických ochorení, ako aj mieru obmedzení realizácie bežných denných aktivít u osôb vo veku 16 a viac rokov.

Ak sa zameriame na kvalitu vlastného zdravia, potom z grafov č. 14 a 15 môžeme vidieť, že v mladších vekových skupinách jednoznačne dominovala odpoveď signalizujúca veľmi dobré a dobré zdravie. Podiel týchto odpovedí s rastúcim vekom však rýchlo klesal. Dynamika tohto trendu bola o niečo rýchlejšia na strane žien. Preto napríklad vo veku 50 – 54 rokov reflektovala len približne polovica opýtaných, že ich zdravotný stav je veľmi dobrý alebo dobrý, kým u rovnako starých mužov to bolo ešte viac ako 60 % respondentov. Prevalu mužov s veľmi dobrým alebo dobrým zdravím pritom identifikujeme ešte vo veku 55 – 59 rokov.

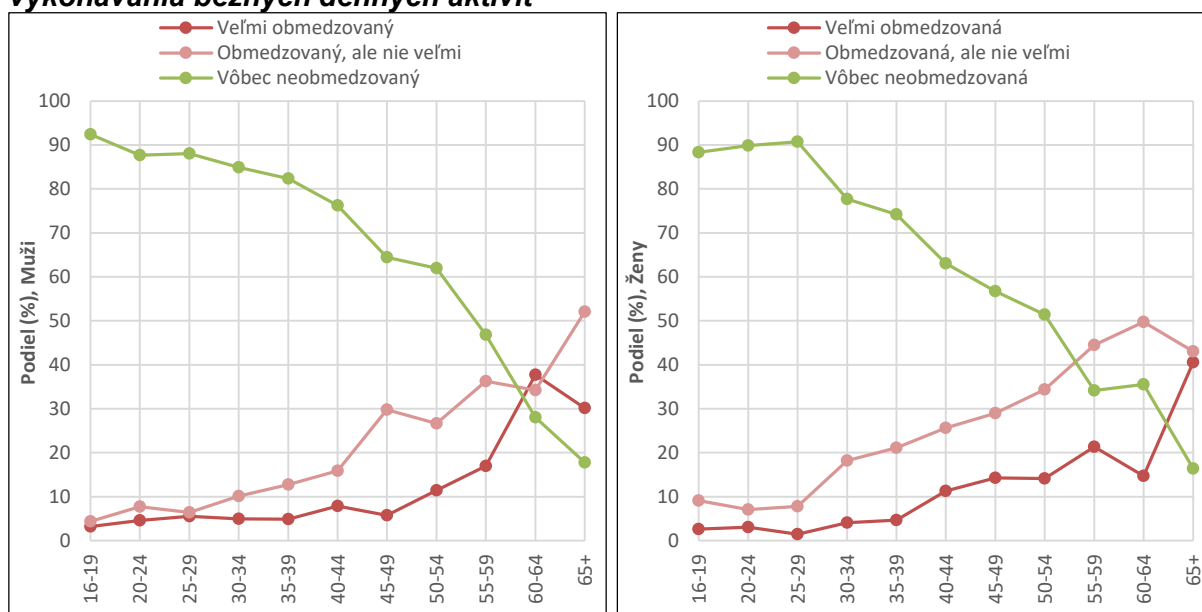
Ako sme už naznačili, s rastúcim vekom sa zvyšuje aj podiel mužov a žien s horšou kvalitou zdravotného stavu. Napríklad vo veku 55 – 59 rokov približne tretina mužov z MRK deklarovala ani dobré, ani zlé zdravie a asi pätina uvádzala zlé a veľmi zlé zdravie. U žien ani dobré, ani zlé zdravie v tomto vekovom intervale uvádzalo približne 37 % respondentiek. Viac ako štvrtina z nich však uviedla zlé a veľmi zlé zdravie. Prevaha tejto skupiny vznikla až v seniorskom veku. U mužov však aj v tomto vekovom spektre sú o niečo častejšie zastúpené osoby s ani dobrým, ani zlým zdravím.

Graf č. 14 a 15: Štruktúra mužov a žien z MRK podľa veku a subjektívneho hodnotenia kvality vlastného zdravia



Zdroj: EU-SILC MRK 2020; výpočty autora

Graf č. 16 a 17: Štruktúra mužov a žien z MRK podľa veku a miery obmedzovania vykonávania bežných denných aktivít



Zdroj: EU-SILC MRK 2020; výpočty autora

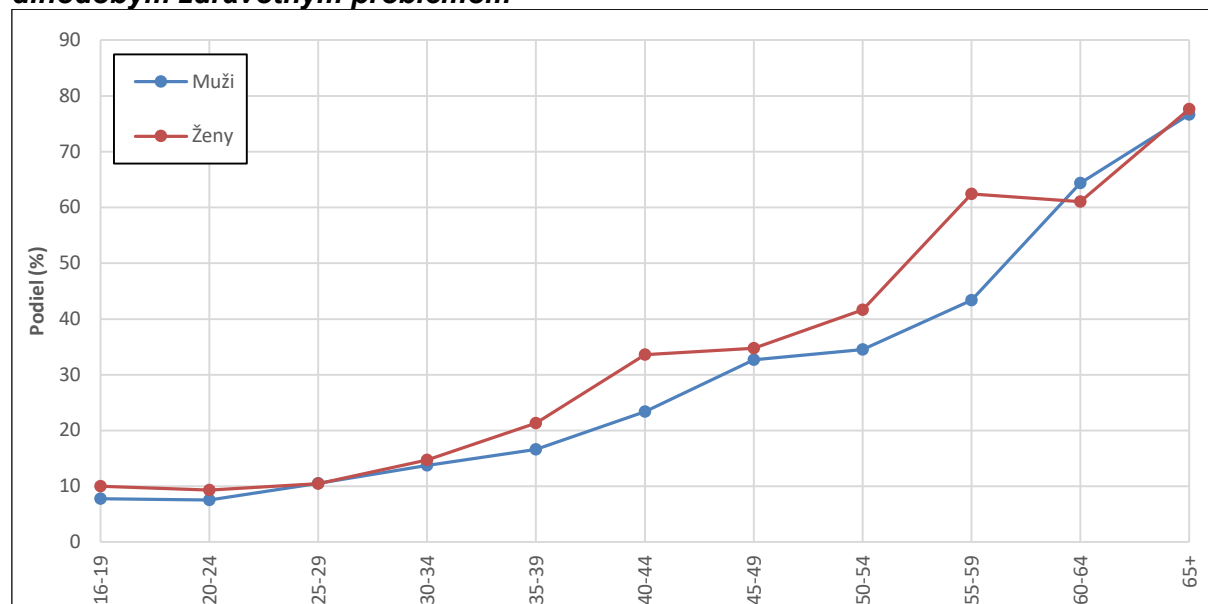
S rastúcim vekom sa u oboch pohlaví znižoval tiež podiel osôb, ktoré v analyzovanom zisťovaní uvádzali, že nie sú vôbec obmedzované vo vykonávaní svojich bežných denných aktivít. Do veku 30 rokov to u oboch pohlaví bolo približne 90 % respondentov. Už vo veku 45 – 49 rokov však nebolo vôbec obmedzovaných len približne 64 % mužov a 57 % žien (graf č. 16 a 17). V najstarších vekoch preto identifikujeme prevahu osôb s nejakým stupňom obmedzenia. Aj v tomto prípade sa ukazuje, že v prostredí MRK reflektujú horšiu kvalitu zdravia ženy, resp. sa miera ich obmedzenia s vekom dynamickejšie zvyšuje. Preto napríklad vo veku 55 – 59 rokov už mali prevahu osoby s obmedzením ale nie veľkým (graf č. 17). Na druhej strane je

však zrejme, že najhorší kvalitatívny stupeň neprevláda (okrem mužov vo veku 60 – 64 rokov) ani v jednej vekovej skupine.

Podľa výsledkov EU-SILC MRK 2020 približne desatina mužov a žien vo veku do 30 rokov uvádza nejaký dlhodobý zdravotný problém alebo dlhodobé ochorenie (graf č. 18). S rastúcim vekom identifikujeme zvyšovanie zastúpenia týchto osôb, pričom opätovne o niečo dynamickejší bol tento trend na strane žien. Vo veku 40 – 44 rokov už približne tretina z nich odpovedala kladne na prítomnosť chronického ochorenia alebo dlhodobého problému a vo veku 50 – 54 rokov to bolo viac ako 40 % opýtaných (graf č. 18).

Najväčší nárast však u žien nachádzame medzi 50. – 54. a 55. – 59. rokom a u mužov medzi vekovými skupinami 55 – 59 a 60 – 64 rokov. V oboch prípadoch nastáva prevaha osôb s dlhodobými zdravotnými problémami alebo ochorením. V seniorskom veku už do tejto skupiny patrí takmer 80 % všetkých opýtaných. Súčasne tiež dochádza k vyrovnaniu diferencií medzi pohlaviami (graf č. 18).

Graf č. 18: Podiel mužov a žien z MRK podľa veku s dlhodobým ochorením alebo dlhodobým zdravotným problémom



Zdroj: EU-SILC MRK 2020; výpočty autora

6. ZÁVER

Údaje získané prostredníctvom zisťovania EU-SILC MRK 2020 predstavujú dôležitú sondy do životných podmienok osôb žijúcich v špecifickom prostredí MRK na Slovensku. Rovnako tiež umožňujú analyzovať niektoré otázky spojené s procesom vzdelávania, vzdelanostnou štruktúrou, ekonomickým postavením, či hodnotením niektorých aspektov vlastného zdravotného stavu.

Získané výsledky potvrdzujú, že typickým znakom vzdelanostnej dráhy mužov i žien z MRK je ich skoré ukončovanie a dosahovanie predovšetkým základného vzdelania 2. stupňa. Navyše nezanedbateľná časť z týchto osôb nadobúda toto vzdelanie v špeciálnych triedach. Len malá časť dospelých pokračuje na strednej škole, a ak sa tak stane, najčastejšie ide o strednú odbornú školu s výučným listom. Vyššie vzdelanie je v podstate marginálnym javom. S tým úzko súvisia aj niektoré syntetické

ukazovatele vzdelanostnej štruktúry, ako je vzdelanostný potenciál a priemerná dĺžka vzdelávania. Oba v prostredí MRK dosahujú veľmi nízke hodnoty, pričom navyše s rastúcim vekom sa situácia ešte zhoršuje.

So skorým ukončovaním vzdelanostnej dráhy v tomto prostredí je úzko prepojený aj veľmi nízky podiel študentov a študentiek vo veku 20 – 24 rokov, vysoká nezamestnanosť v tomto vekovom spektre a u žien pomerne vysoké podiely osôb v pozícii inej neaktívnej osoby (najčastejšie na rodičovskej dovolenke), v domácnosti a pod. Celkovo výsledky zisťovania EU-SILC MRK 2020 tiež potvrdili, že muži a najmä ženy z tohto prostredia sú poznačení vysokou nezamestnanosťou a v poreprodukčnom veku aj častejšou prítomnosťou invalidných osôb. Nie je preto prekvapujúce, že tieto osoby len v minimálnej miere zostávajú na pracovnom trhu po dosiahnutí 65 rokov.

Z hľadiska zdravotného stavu výsledky jednoznačne ukazujú na jeho zhoršovanie s vekom, pričom dynamickejšie tento trend prebieha u žien. Potvrdzuje sa to nielen v kontexte subjektívneho hodnotenia kvality zdravia, ale aj v miere obmedzovania bežných činností či prítomnosti chronických ochorení a dlhodobých zdravotných problémov.

LITERATÚRA

- [1] KALIBOVÁ, K.: Špecifické rysy úmrtnosti romskej populácie v Československu. In: ÚMRTNOST a stárnutí obyvateľstva v ČSSR. Acta Demographica VIII. Praha: Československá demografická spoločnosť pri ČSAV, 1988, s. 63 – 70.
- [2] KALIBOVÁ, K.: Charakteristika úmrtnostných pomerů romskej populácie v ČSSR. Demografie, 1989, č. 3, s. 239 – 250.
- [3] KURNIENKOVÁ, N.: Zdravotná starostlivosť v sociálne vylúčených rómskych komunitách. Madrid: Fundación Secretariado Gitano, 2007. 68 s.
- [4] LORAN, T. Sociálna politika a zamestnanosť Rómov. In: Vašečka, M. (ed.) Čačipen Pal o Roma Súhrnná správa o Rómoch na Slovensku. Bratislava: Inštitút pre verejné otázky, 2002, s. 565 – 586.
- [5] MARKOVIČ, F. – PLACHÁ, Ľ.: Príjmy a životné podmienky v marginalizovaných rómskych komunitách: Vybrané ukazovatele zo zisťovania EU SILK_MRK 2020. Bratislava: Úrad splnomocnenca vlády SR pre rómske komunity, 2021.
- [6] MAZOUCH, P. – FISCHER, J.: Lidský kapitál. Měření, souvislosti a prognózy. Praha: C. H. Beck, 2011.
- [7] POPPER, M. – SZEGHY, P. – ŠARKOZY, Š.: Rómska populácia a zdravie: Analýza situácie na Slovensku. Bratislava: Partners for Democratic Change Slovakia, 2009. 95 s.
- [8] RADIČOVÁ, I.: Chudoba Rómov vo vzťahu k trhu práce v Slovenskej republike. In: Sociológia, 2001, č. 5, s. 439 – 456.
- [9] RADIČOVÁ, I.: Rómovia na prahu transformácie. In: Vašečka M. (ed.) Čačipen Pal o Roma. Súhrnná správa o Rómoch na Slovensku. Bratislava: Inštitút pre verejné otázky, 2002. s. 79 – 92.
- [10] ŠAŠKO, P.: Zdravotná situácia romskej populácie. In: VAŠEČKA, M (ed.) Čačipen Pal o Roma. Súhrnná správa o Rómoch na Slovensku. Bratislava: Inštitút pre verejné otázky, 2002. s. 657 – 678.
- [11] ŠPROCHA, B.: Úmrtnosť a zdravotný stav romskej populácie na Slovensku, 1. časť. Slovenská štatistika a demografia, 2012. č. 3, s. 86 – 101.

[12] ŠPROCHA, B.: Úmrtnosť a zdravotný stav rómskej populácie na Slovensku, 2. časť. Slovenská štatistika a demografia, 2012. č. 3, s. 18 – 32.

[13] ŠPROCHA, B.: Reprodukcia rómskeho obyvateľstva na Slovensku a prognóza jeho populačného vývoja. Bratislava: Prognostický ústav SAV, 2014.

[14] ŠPROCHA, B. – ĎURČEK, P.: Rómovia na Slovensku v sčítaniach obyvateľov 1980 – 2011. Bratislava: INFOSTAT, 2017.

RESUMÉ

Zisťovanie EU-SILC MRK 2020 prináša dôležité a v mnohých smeroch jedinečné údaje o životných podmienkach osôb žijúcich v marginalizovaných rómskych komunitách na Slovensku. Keďže obsahuje aj rôzne štrukturálne znaky, umožňuje hodnotiť aj vybrané aspekty socio-ekonomických, kultúrnych a demografických štruktúr. Cieľom tohto príspevku bolo analyzovať vzdelanostnú štruktúru, ekonomické postavenie mužov a žien a ich zdravotný stav.

Získané výsledky potvrdili skoré ukončovanie vzdelanostnej dráhy. Preto väčšina osôb žijúca v prostredí marginalizovaných rómskych komunít na Slovensku má len základné vzdelanie. Aj to sa pomerne často nadobúda v špecializovaných školách. Podiel osôb s vyšším vzdelaním je nízky a väčšinou ide len o osoby so stredoškolským (učňovským) vzdelaním bez maturity. S tým súvisí veľmi nízka priemerná dĺžka vzdelávania, ako aj vzdelanostný potenciál. Navyše tie sa v mladších generáciách výraznejšie neodlišujú od situácie u starších osôb. So skorším ukončovaním vzdelávania súvisia aj veľmi nízke podiely študujúcich osôb a naopak vysoká nezamestnanosť už v mladom veku. Práve dlhodobá nezamestnanosť a vysoký podiel nezamestnaných mužov a žien predstavujú hlavnú črtu ekonomického postavenia osôb v tomto špecifickom prostredí. Najmä u žien identifikujeme tiež nezanedbateľný podiel osôb v domácnosti a na rodičovskej alebo materskej dovolenke. S rastúcim vekom sa postupne tiež zvyšuje zastúpenie osôb na invalidnom dôchodku. To môže byť spojené so zhoršujúcim sa zdravotným stavom, ktorý deklarujú obe pohlavia. O niečo dynamickejšie sa však kvalita zdravia podľa odpovedí zhoršuje u žien. Zvyšuje sa tiež podiel osôb s nejakou mierou obmedzenia bežných denných aktivít, ako aj prítomnosťou chronického alebo iného dlhodobo trvajúceho ochorenia.

RESUME

The EU-SILC MRK 2020 survey provide important and in many ways unique data on the living conditions of people living in the marginalized Roma communities in Slovakia. Since it contains various structural features, it also enables the evaluation of selected aspects of socio-economic, cultural and demographic structures.

The aim of this contribution was to analyze the educational structure, the economic status of men and women and their health status. The obtained results confirmed the early termination of educational paths. Therefore, the majority of people living in the environment of marginalized Roma communities in Slovakia have only basic education. This is also quite often acquired in specialized schools. The share of people with a higher education is low, and mostly, moreover, it is only people with a secondary school (apprenticeship) education without a high school diploma. This is related to the very low average length of education, as well as the educational potential. Moreover, these do not significantly differ from older people in the younger generations. The early completion of education is also related to very low proportions of people studying and, on the contrary, high unemployment already at a young age. Long-term unemployment and a high proportion of unemployed men and women represent the main feature of the economic status of people in this specific environment. Especially among

women, we also identify a significant share of people in the household and on parental or maternity leave. With increasing age, the representation of persons on disability pension also gradually increases. This may be associated with deteriorating health status, which is reported by both sexes. However, according to the responses, the quality of health deteriorates somewhat more dynamically among women. The share of people with some degree of restriction of normal daily activities, as well as the presence of a chronic or other protracted illness, is also increasing.

PROFESIJNÝ ŽIVOTOPIS

Doc. RNDr. PhDr. 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 INFOSTAT-e 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 vplyvu na spoločnosť. Okrem toho sa zameriava na analýzu vybraných populačných štruktúr, reprodukčného správania rómskeho obyvateľstva na Slovensku a na otázky konštrukcie populačných prognóz.

KONTAKT

branislav.sprocha@gmail.com

Roman PAVELKA
Štatistický úrad Slovenskej republiky

IMPUTACE CHYBĚJÍCÍCH DAT POMOCÍ BAYESOVSKÉHO MODELOVÁNÍ

MISSING DATA IMPUTATION USING BAYESIAN DATA MODELLING

ABSTRAKT

Proč jsou chybějící údaje problém? Protože běžné statistické metody a software předpokládají, že všechny hodnoty u všech proměnných v matici dat jsou pozorovány za všechny jednotky účastné ve statistickém zjišťování. Výchozí metodou řešení neodpovědí prakticky u všech statistických softwarů je prosté vymazání případů s chybějícími údaji ukazatelů, které jsou předmětem zájmu. Nejzřetelnější nevýhodou vymazání seznamu (jednotek) je to, že se často vymaže velká část vzorku sesbíraných statistických dat. Odstranění sesbíraných dat, které nejsou vhodné k dalšímu statistickému zpracování, může vést k vážné ztrátě statistické síly analýz. Výzkumníci pochopitelně neradi vyřazují údaje, jejichž sběru věnovali mnoho času, peněz a úsilí, a proto se staly populárními různé metody „záchrany“ případů s chybějícími údaji. Moderní metodou pro doplnění neúplných dat se v posledních několika desetiletích stává bayesovská inference. Bayesovská pravděpodobnost a statistika je mnohem více než všeobecně známý Bayesův vzorec a jeho občasné použití v ukázkových či ilustrativně orientovaných příkladech při výkladu operací s pravděpodobnostmi náhodných jevů. Bayesův pravděpodobnostní vzorec (nazývaný také zákon o inverzní pravděpodobnosti) se především používá v souvislosti s úsudky o neznámém modelu na základě známých dat. Toto dává možnosti pro použití Bayesova vzorce při imputování nepozorovaných dat (neznámý model) na základě dat zjištěných.

ABSTRACT

Why are missing data a problem? Because common statistical methods and software assume that all values for all variables in the data matrix are observed for all units participating in the statistical survey. The default method of dealing with nonresponse in all statistical software is to simply delete cases with missing data for the indicators of interest. The most obvious disadvantage of list (unit) removal is that it often deletes a large portion of the sample of collected statistical data. Removal of the collected data not suitable for further statistical processing can lead to a serious loss of statistical power of the analyses. Researchers are understandably reluctant to discard data that they have spent a lot of time, money and effort collecting, so various methods of 'rescuing' cases with missing data have become popular. The Bayesian inference has become a modern method for completing incomplete data over the last few decades. The Bayesian probability and statistics is far more than the well-known Bayesian formula and its occasional use in demonstrative or illustrative examples in explaining operations with probabilities of random phenomena. The Bayesian' probability formula (also called the law of inverse probability) is primarily used in the context of making judgments about an unknown model based on known data. This provides opportunities for using the Bayesian' formula in imputing unobserved data (unknown model) based on observed data.

KLÍČOVÁ SLOVA

bayesovská inference, inverzní pravděpodobnost, mechanismus chybění, imputace dat

KEY WORDS

Bayesian inference, inverse probability, mechanism of missingness, data imputation

1. ÚVOD DO PROBLEMATIKY CHYBĚJÍCÍCH ÚDAJŮ

Problém neúplných dat se v praxi vyskytuje velmi často. Například míra návratnosti vyplněných statistických formulářů v podnikových šetřeních se pohybuje v závislosti na typu, periodicitě a složitosti dotazování od 50 % do 70 %, přičemž odpovědi vybraných statistických jednotek mnohdy nebývají kompletní a také ani bezchybné. Vyšší mírou návratnosti se vyznačují sociální statistiky, například šetření domácností, kde vybrané domácnosti jsou dotazovány pomocí speciálně školených dotazovatelů. Významným faktorem při zjišťování potřebných údajů je neochota respondentů, kteří často odpověď neznají, odpovědi si nepamatují anebo odpovědět jednoduše nechťejí. Následná analýza takto pořízených dat často vychází z předpokladu, že proces, který zapříčinil chybějící data, lze při statistické inferenci více či méně ignorovat. Zde je však potřebné si odpovědět na otázku: jedná se o správné statistické postupy?

Neúplné, resp. chybějící informace také narušují zjišťované údaje, vnášejí do nich prvek zkreslení, znehodnocují výsledky a závěry, činí je nevhodnými pro analýzy standardními statistickými metodami, a způsobují, že taková data úřady mohou kvůli odchylkám odmítnout. Náprava takto zjišťovaných dat nespočívá pouze v zanedbání jednotek s neúplnými či chybějícími informacemi. Pokud se tyto jednotky s chybějícími údaji jednoduše vyloučí, ovlivní to zejména statistickou sílu realizovaného zjišťování. Zároveň je pravděpodobné, že zpravodajské jednotky s nevyplněnými hodnotami mohou být jednotkami s extrémními anebo odlehlými hodnotami. Vyloučení těchto jednotek povede k podhodnocení variability, a tedy k zúžení intervalu spolehlivosti. Chybějící údaje samy o sobě představují různé problémy, například:

- absence údajů snižuje statistickou sílu, která se vztahuje k pravděpodobnosti, že test zamítne nulovou hypotézu, pokud je nepravdivá,
- ztracené údaje mohou způsobit zkreslení odhadu parametrů,
- chybějící údaje mohou snížit reprezentativnost vzorku a
- může zkomplikovat analýzu zjišťování nebo studie.

Jako chybějící údaje jsou považovány údaje, které v datové matici sledovaných hodnot chybí u některých (případně u všech) proměnných a u některých (případně u všech) respondentů (případů, resp. zpravodajských jednotek). Tyto chybějící údaje z datové matice jsou problematické především z toho důvodu, že většina současných analytických programů a nástrojů předpokládá datovou matici úplnou. Proto může být tento postup zpracování dat nevhodný a výsledné statistické úsudky mohou být více či méně chybné.

2. POUŽITÉ NÁHODNÉ VELIČINY A JEJICH ZNAČENÍ

Na rozdíl od většiny úloh matematické statistiky, která se soustředí na odhad hypotetických neznámých parametrů (např. průměrné hodnoty hypotetického normálního rozdělení, které vygenerovala data), metodika výběrových šetření se soustředí na odhad neznámých pozorovatelných veličin, jako je např. průměrný příjem rodin v určité konečné populaci. Výraz parametr slouží k označení hypotetických neznámých a nepoznatelných údajů. Nepozorované hodnoty v konečné populaci se považují za chybějící údaje, tj. chybějící hodnoty pozorovatelných veličin. Pozorovatelné veličiny, které jsou předmětem zájmu v konečné populaci, budou pro

účely statistického šetření uspořádány do jednotek (řádky) \times proměnné (sloupce). I když snahou v každém statistickém šetření je sledovat všechny hodnoty v této matici, zpravidla v každém šetření populace nejsou všechny hodnoty matice pozorovány.

V konečné populaci složené z N jednotek jsou zpravidla definovány dva druhy proměnných. První druh zahrnuje proměnné, které popisují charakteristiky jednotek zajímavé pro účely statistického zjišťování: a to plně pozorovatelné vysvětlující proměnné (kovariáty) X a vysvětlované (závislé na vysvětlujících proměnných) proměnné Y , které jsou předmětem výzkumného zájmu. Druhý druh proměnných zahrnuje indikátorové proměnné – indikátory výběru I a indikátory neodpovědí R , které jsou potřebné k popisu, které hodnoty jsou pozorované a které nepozorované. Indikátorové proměnné I a R také pomáhají vyvodit správné induktivní závěry o parametrech populace. Matici proměnných v konečné populaci zobrazuje tabulka č. 1.

Tabulka č. 1: Matice proměnných v konečné populaci o N jednotkách

	Kovariáty X			Závisle proměnné Y			Indikátory výběru I			Indikátory odpovědí R			
	1	\dots	q	1	\dots	p	1	\dots	p	1	\dots	p	
Statistické jednotky v konečné populaci	1	X_{11}	\dots	X_{1q}	Y_{11}	\dots	Y_{1p}	I_{11}	\dots	I_{1p}	R_{11}	\dots	R_{1p}
	\vdots	\vdots		\vdots	\vdots		\vdots	\vdots		\vdots	\vdots	\vdots	
	i	X_{i1}	\dots	X_{iq}	Y_{i1}	\dots	Y_{ip}	I_{i1}	\dots	I_{ip}	R_{i1}	\dots	R_{ip}
	\vdots	\vdots		\vdots	\vdots		\vdots	\vdots		\vdots	\vdots	\vdots	
	N	X_{N1}	\dots	X_{Nq}	Y_{N1}	\dots	Y_{Np}	I_{N1}	\dots	I_{Np}	R_{N1}	\dots	R_{Np}

Zdroj: [1, s. 29]

X jsou plně pozorované kovariáty, jako jsou indikátor straty nebo velikost jednotky, a jsou zaznamenány pro všechny jednotky N v populaci. Existuje-li více než 1 plně pozorovatelná kovariáta, bude X_i řádkový vektor. Y vyjadřují proměnné, jejichž hodnoty nejsou známy pro všechny jednotky v populaci. Existuje-li více než 1 závislá proměnná, bude Y_i řádkový vektor. Obvykle bude Y zahrnovat pouze ty proměnné, které jsou primárně předmětem zájmu šetření, například jako je příjem jednotlivců, objemy tržeb, obratu nebo přidané hodnoty. Indikátor I bude ukazatelem pro zahrnutí/nezahrnutí do šetření. V případě pouze jedné závislé proměnné Y_i je ukazatel I binární, přičemž $I_i = 1$ značí že i -tá jednotka je do šetření zahrnuta (tj. byl učiněn pokus o zaznamenání proměnné Y_i) a $I_i = 0$ je i -tá jednotka ze šetření vyloučena (tj. nebyl učiněn žádný pokus o zaznamenání Y_i). Reprezentuje-li Y_i více než 1 proměnnou, např. p proměnných, I_i je také vektor s p prvky; prvek I_{ij} indikuje, zda byl učiněn pokus o zaznamenání hodnoty Y_{ij} ($I_{ij} = 1$) nebo nebyl ($I_{ij} = 0$). Předpokládá se, že matice indikátorů je plně pozorovaná. Indikátor odpovědí R je indikátorem pro odpověď či neodpověď jednotky. V případě pouze 1 závislé proměnné, je indikátorová proměnná R_i binární, přičemž $R_i = 1$ indikuje že i -tá jednotka zaznamenala hodnotu závislé proměnné (tj. hodnota Y_i je pozorována, pokud je učiněn pokus o zaznamenání odpovědi Y_i v případě, že i -tá jednotka bude zahrnuta do šetření) a $R_i = 0$ indikuje, že hodnota závislé proměnné i -té jednotky pozorována není (samozřejmě v případě zahrnutí i -té jednotky do šetření). Pokud Y_i vektor závislých proměnných obsahuje p proměnných, R_i je také vektorem s p proměnnými, kde prvek R_{ij} indikuje odpověď či neodpověď proměnné Y_{ij} . Předpokládá se, že indikátor R_{ij} je znám, je-li indikátor zahrnutí $I_{ij} = 1$. Je-li indikátor zahrnutí i -té jednotky

$I_i = 0$, je indikátor odpovědi R_i u i -té jednotky nedefinován. Ačkoliv výběrová šetření mohou obecně být víceúrovňová se složitějšími návrhy výběru, bude pro účely tohoto článku prezentováno výběrové šetření jednostupňové s jednoduchým náhodným výběrem. V případě ostatních typů statistických šetření se následující matematické vztahy musí příslušně upravit.

3. PRAVDĚPODOBNOSTNÍ SPECIFIKACE POUŽITÝCH PROMĚNNÝCH

Pravděpodobnostní specifikace pro indikátorové proměnné I a R jsou potřebné k tomu, aby bylo možné vyvození správných závěrů pro neznámé hodnoty v konečné populaci. Tyto pravděpodobnostní modely se nazývají „mechanismy“, a nikoliv „modely“, aby mohly být odlišeny od obvyklejších statistických modelů týkajících se rozdělení proměnných X, Y [12].

Mechanismus výběru je podle [12] specifikován jako podmíněná pravděpodobnost daná výrazem $Pr(I|X, Y, R)$. Všechny formální statistické metody, které se používají k vyvozování závěrů o populaci, vyžadují explicitní nebo implicitní popis mechanismu výběru. Mechanismus výběru je považován za pravděpodobnostní, pokud každé hodnotě závislé proměnné Y_{ij} přiřazuje kladnou hodnotu pravděpodobnosti, že bude zahrnuta do výběrového souboru, tj.

$$Pr(I_{ij} = 1|X, Y, R) > 0 \quad \text{pro všechny } I_{ij}, \quad (1)$$

Je-li pro j -tou hodnotu závislé proměnné Y i -té jednotky pravděpodobnost zahrnutí do výběrového souboru rovná 0, tj. $Pr(I_{ij} = 1|X, Y, R) = 0$, je mechanismus výběru považován jako nepravděpodobnostní. Standardní výběrové techniky jsou navrženy tak, aby umožňovaly jednu obecně přijatelnou specifikaci pro $Pr(I_{ij} = 1|X, Y, R)$. Často se jedná o pravděpodobnostní mechanismus výběru, jak je definován v (1) a (2).

$$Pr(I|X, Y, R) = Pr(I|X) \quad \text{pro všechny kombinace } (I, X, Y, R). \quad (2)$$

Z formulací (1) a (2) jednoduše vyplývá, že:

- mechanismus výběru musí zajistit, aby každá hodnota Y z populace měla šanci být zahrnuta do výběrového souboru (viz 1) a
- mechanismus výběru vzorků může k výběru použít pozorované proměnné X , jako jsou ukazatele strat, velikosti, apod. (viz 2).

Pro techniky výběru ale nemohou být použity hodnoty Y nebo R , protože nejsou v okamžiku výběru známé.

Příkladem mechanismu výběru nepravděpodobnostního šetření je telefonický průzkum v populaci, která však zahrnuje i některé domácnosti bez telefonů. Pokud proměnná X_i je počet telefonů v i -té domácnosti a vyšetřeny mají být všechny jednotky, mechanismus výběru je nepravděpodobnostní a je daný rovnicí

$$Pr(I|X, Y, R) = \begin{cases} 1 & \text{if } I = 1 \text{ když } X_i > 0 \text{ a } I = 0 \text{ když } X_i = 0 \text{ a} \\ 0 & \text{jinak} \end{cases}. \quad (3)$$

Mechanismus response dat lze podobně jako v předešlém případě specifikovat jako podmíněnou pravděpodobnost. Na rozdíl od mechanismu výběru, mechanismus odpovědi nezávisí na způsobu výběru respondentů. Proto lze tento mechanismus vyjádřit jako pravděpodobnost podmíněnou vysvětlovanými anebo vysvětlujícími proměnnými, tj. $Pr(\mathbf{R}|\mathbf{X}, \mathbf{Y})$. Pokud každé hodnotě závislé proměnné Y_{ij} přiřazuje mechanismus odpovědi kladnou hodnotu pravděpodobnosti, že bude hodnota proměnné Y_{ij} pozorovaná, jedná se o pravděpodobnostní mechanismus odpovědi, tj.

$$Pr(R_{ij} = 1|\mathbf{X}, \mathbf{Y}) > 0 \quad \text{pro všechny } i, j. \quad (4)$$

Předpoklady o specifických formách mechanismu odpovědi mohou mít zásadní význam pro vhodné úpravy při vyvozování závěrů o neodpovědích.

Jelikož dvojice proměnných (\mathbf{X}, \mathbf{Y}) je matice náhodných proměnných (viz tabulka č. 1) o rozměrech $N \times (p + q)$, je nutné specifikovat sdružené rozdělení $Pr(\mathbf{X}, \mathbf{Y})$. Nechť pro (\mathbf{X}, \mathbf{Y}) existuje rozdělení $f(\mathbf{X}, \mathbf{Y})$, jehož předpokladem je vzájemná zaměnitelnost $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, \dots, N$ tj. řádků matice z tabulky č. 1¹. Potom podle de Finettiho věty [3] vyplývá, že $Pr(\mathbf{X}, \mathbf{Y})$ lze zapsat ve tvaru, kde při daném vektoru parametrů θ s marginální (apriorní) hustotou $Pr(\theta)$, jsou $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, \dots, N$, nezávisle a identicky rozděleny se společným rozdělením $f_{YX}(\mathbf{X}_i, \mathbf{Y}_i|\theta)$. Potom podle de Finettiho věty lze pravděpodobnost $Pr(\mathbf{X}, \mathbf{Y})$ zapsat ve tvaru:

$$Pr(\mathbf{X}, \mathbf{Y}) = \int_{\theta} [\prod_{i=1}^N f_{YX}(\mathbf{X}_i, \mathbf{Y}_i|\theta)] Pr(\theta) d\theta. \quad (5)$$

Konkrétním příkladem použití (5) je jednoduchý normální model, ve kterém vystupuje závislá proměnná \mathbf{Y} jako skalár (vektor s jedinou složkou, tj. $\mathbf{Y} = (Y_1, \dots, Y_N)^T$) bez vektoru vysvětlovaných proměnných \mathbf{X} . Pro každé $i = 1, \dots, N$, je rozdělení vysvětlované proměnné identické a navzájem nezávislé rozdělení odpovídající normálnímu rozdělení $N(\mu, \sigma^2)$, každé s vektorem parametrů $\theta = (\mu, \sigma^2)$ a rozdělením θ úměrným σ^{-2} , tj.

$$Pr((Y_1, \dots, Y_N)^T|\mu, \sigma^2) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(Y_i - \mu)^2}{2\sigma^2}\right] \quad (6)$$

a

$$Pr(\mu, \sigma^2) \propto \sigma^{-2}. \quad (7)$$

Potom podmíněné rozdělení parametru μ za podmínky realizace parametru σ^2 a proměnné \mathbf{Y} , tj. $(\sigma^2, y_1, \dots, y_n)$, je $N(\bar{y}, \sigma^2/n)$, kde $\bar{y} = \sum_{i=1}^n y_i/n$, tj.

$$Pr(\mu|\sigma^2, (y_1, \dots, y_n)^T) = \sqrt{\frac{n}{2\pi\sigma^2}} \exp\left[-n \frac{(\mu - \bar{y})^2}{2\sigma^2}\right]. \quad (8)$$

Podmíněné rozdělení parametru σ^2 za podmínky realizace proměnné $(Y_1, \dots, Y_n)^T$ je $(n-1)s^2\chi_{n-1}^{-2}$, kde $s^2 = \sum_1^n (y - \bar{y})^2/(n-1)$ a χ_{n-1}^{-2} je inverze rozdělení χ^2 s $(n-1)$ stupni volnosti.

¹ Předpokladem je náhodné přiřazení indexů jednotek, které musí být v (řádkových) indexech jednotek zaměnitelné a se stejnou pravděpodobností: $Pr(\mathbf{X}, \mathbf{Y}) = Pr(\text{row} - \text{perm}(\mathbf{X}, \mathbf{Y}))$, kde $\text{row} - \text{perm}(\mathbf{X}, \mathbf{Y})$ je libovolná permutace řádků (\mathbf{X}, \mathbf{Y}) .

Pro podmíněné rozdělení vektoru parametrů (μ, σ^2) za podmínky realizace proměnných $(Y_1, \dots, Y_n)^T$ potom platí

$$Pr(\mu, \sigma^2 | (y_1, \dots, y_n)^T) \propto Pr((Y_1, \dots, Y_n)^T | \mu, \sigma^2) Pr(\mu, \sigma^2). \quad (9)$$

4. BAYESOVSKÁ INFERENCE V PODMÍNKÁCH CHYBĚJÍCÍCH HODNOT

Nechť platí pro $inc = \{(i, j) | I_{ij} = 1\}$ tak, že Y_{inc} označuje složky proměnné Y zahrnuté do výběrového souboru a R_{inc} označuje složky vektoru R zahrnuté do stejného souboru; analogicky je definován výraz $exc = \{(i, j) | I_{ij} = 0\}$. Pro Y a R potom platí $Y = (Y_{inc}, Y_{exc})^T$ a $R = (R_{inc}, R_{exc})^T$. Jak Y_{exc} a R_{exc} jsou nepozorované, protože nejsou do výběrového souboru zahrnuty; R_{inc} je plně pozorována, ale Y_{inc} je plně pozorována pouze za podmínky neexistence neodpovědí.

Nechť platí pro $obs = \{(i, j) | I_{ij} = 1 \text{ a } R_{ij} = 1\}$ tak, že Y_{obs} označuje složky proměnné Y pozorované ve výběrovém souboru; každá složka má přiřazen indikátor $R_{ij} = 1$. Analogicky výraz $mis = \{(i, j) | I_{ij} = 1 \text{ a } R_{ij} = 0\}$ bude použit pro komponenty proměnné Y , které jsou do výběrového souboru vybrané, ale nepozorované (chybějící) s přiřazenou hodnotou $R_{ij} = 0$, tj. Y_{mis} a platí, že $Y_{inc} = (Y_{obs}, Y_{mis})^T$. Dále nechť $nob = \{(i, j) | I_{ij} = 0 \text{ anebo } R_{ij} = 0\}$, kde $Y_{nob} = (Y_{exc}, Y_{mis})^T$ vyjadřuje nepozorované složky vysvětlované proměnné Y . Na základě tohoto značení lze složky proměnné Y rozdělit na složky pozorované Y_{obs} a složky nepozorované Y_{nob} , tj. $Y = (Y_{obs}, Y_{nob})^T$.

Bayesovská inference pro populační parametr $Q(X, Y)$ se odvozuje z jeho aposteriorního rozdělení, tj. jeho podmíněného rozdělení vzhledem k pozorovaným hodnotám X, Y_{obs}, R_{inc}, I vypočteným podle zadaných modelů. Použije-li se výše zavedená notace, lze toto aposteriorní rozdělení zapsat jako $Pr(Q | X, Y_{obs}, R_{inc}, I)$. Je-li například aposteriorní rozdělení skaláru Q normální o průměru $\hat{Q} = \hat{Q}(X, Y_{obs}, R_{inc}, I)$ a rozptylem $\hat{U} = \hat{U}(X, Y_{obs}, R_{inc}, I)$, potom interval, který pokrývá populační parametr Q s pravděpodobností 95%, je $\hat{Q} \pm 1,96\hat{U}^{1/2}$.

Bayesovská inference také umožňuje vypočítat aposteriorní rozdělení pro nepozorované (chybějící) hodnoty Y_{nob} . Jelikož $Q(X, Y)$ je funkcí pozorovaných hodnot v X, Y_{obs} a nepozorovaných (chybějících) hodnotách Y_{nob} , lze aposteriorní rozdělení nepozorovaných (chybějících) hodnot Y_{nob} vyjádřit jako:

$$Pr(Q | X, Y_{obs}, R_{inc}, I) = \int_{\omega(Q)} Pr(Y_{nob} | X, Y_{obs}, R_{inc}, I) dY_{nob}, \quad (10)$$

kde

$$\omega(Q) = \{Y_{nob} | Q(X, Y)\}. \quad (11)$$

Aposteriorní rozdělení nepozorovaných (chybějících) hodnot Y_{nob} lze přímo vyjádřit pomocí specifikace pro proměnné (X, Y) za pomoci mechanismů výběru a odpovědi jako

$$Pr(\mathbf{Y}_{nob}|\mathbf{X}, \mathbf{Y}_{obs}, \mathbf{R}_{inc}, \mathbf{I}) = \frac{\int_{\mathbf{R}_{exc}} Pr(\mathbf{X}, \mathbf{Y})Pr(\mathbf{I}|\mathbf{X}, \mathbf{Y}, \mathbf{R})Pr(\mathbf{R}|\mathbf{X}, \mathbf{Y})d\mathbf{R}_{exc}}{\int_{\mathbf{Y}_{nob}} \int_{\mathbf{R}_{exc}} Pr(\mathbf{X}, \mathbf{Y})Pr(\mathbf{I}|\mathbf{X}, \mathbf{Y}, \mathbf{R})Pr(\mathbf{R}|\mathbf{X}, \mathbf{Y})d\mathbf{R}_{exc}d\mathbf{Y}_{nob}} \quad (12)$$

kde pravděpodobnosti jsou vyhodnocovány v pozorovaných hodnotách (realizaci) $(\mathbf{Y}_{nob}|\mathbf{X}, \mathbf{Y}_{obs}, \mathbf{R}_{inc}, \mathbf{I})$.

5. MECHANISMUS TVORBY CHYBĚJÍCÍCH DAT A JEHO IGNOROVATELNOST

Jako příčinu vzniku procesu, který generuje chybějící údaje ve statistickém šetření, lze podle [12] považovat způsob výběru vzorků, tj. mechanismus výběru $Pr(\mathbf{I}|\mathbf{X}, \mathbf{Y}, \mathbf{R})$ a míra dosažených pozorovaných hodnot, tj. mechanismus odpovědí $Pr(\mathbf{R}|\mathbf{X}, \mathbf{Y})$. V rámci statistické inference za podmínek chybějících dat se k odhadům o populaci používá statistický model pro zjišťované proměnné (\mathbf{X}, \mathbf{Y}) vzhledem k danému parametru θ , tj. $f(\mathbf{X}, \mathbf{Y}|\theta)$, doplněný o statistický model pro indikátory chybějících dat podmíněný konkrétní realizací sledovaných náhodných proměnných (\mathbf{X}, \mathbf{Y}) a parametrem mechanismu φ , tj. $g(\mathbf{I}, \mathbf{M}|\mathbf{X}, \mathbf{Y}, \varphi)$. Symbol \mathbf{M} označuje plně pozorovanou matici binárních indikátorů chybějících dat, přičemž $\mathbf{M} = \{M_{ij}\} = \{I_{ij}R_{ij}\}$ s navzájem jednoznačnými (prostými) funkcemi (\mathbf{I}, \mathbf{M}) a $(\mathbf{I}, \mathbf{R}_{inc})$.

Je-li model chybějících dat $g(\mathbf{I}, \mathbf{M}|\mathbf{X}, \mathbf{Y}, \varphi)$ po realizaci pozorovaných hodnot náhodných proměnných $\mathbf{I}, \mathbf{M}, \mathbf{X}$ a hodnot pozorovaných složek proměnné \mathbf{Y} vyhodnocen jako nezávislý na nepozorovaných (chybějících) složkách proměnné \mathbf{Y} , tj. na \mathbf{Y}_{nob} , potom podle [11] chybějící data mohou být považována za náhodně chybějící, označované výrazem *missing-at-random* (MAR), tj.

$$g(\mathbf{I}, \mathbf{M}|\mathbf{X}, \mathbf{Y}, \varphi) = g(\mathbf{I}, \mathbf{M}|\mathbf{X}, \mathbf{Y}_{obs}, \varphi). \quad (13)$$

Pokud jsou navíc parametry θ a φ odlišné, tj. nezávislé, potom Věta 8.1 v [13] dokazuje, že bayesovská inference může zanedbat proces, který generuje chybějící data, a umožňuje odhadnout parametr θ na základě $f(\mathbf{X}, \mathbf{Y}|\theta)$, apriorního rozdělení $Pr(\theta)$ a pozorovaných hodnot $(\mathbf{X}, \mathbf{Y}_{obs})$.

V kontextu bayesovské inference pro náhodnou proměnnou \mathbf{Y} je kombinace náhodně chybějících údajů a rozdílných hodnot parametrů θ a φ prakticky ekvivalentní tomu, že mechanismy výběru a odpovědí jsou ignorovatelné. Podle [12] platí, že pokud jsou θ a φ apriorně nezávislé a chybějící data jsou náhodně chybějící (typu MAR), pak pro mechanismus výběru a mechanismus odpovědí platí, že jsou shodné, tj.

$$Pr(\mathbf{Y}_{nob}|\mathbf{X}, \mathbf{Y}_{obs}, \mathbf{R}_{inc}, \mathbf{I}) = Pr(\mathbf{Y}_{nob}|\mathbf{X}, \mathbf{Y}_{obs}, \mathbf{M}, \mathbf{I}). \quad (14)$$

Praktický význam toho, že mechanismus tvorby chybějících dat je ignorovatelný, spočívá ve skutečnosti, že aposteriorní rozdělení nepozorované části proměnné \mathbf{Y} , tedy \mathbf{Y}_{nob} , může být zjištěno na základě pozorovaných hodnot $\mathbf{X}, \mathbf{Y}_{obs}$ a specifikací konkrétního rozdělení pro náhodné proměnné \mathbf{X}, \mathbf{Y} , tj.:

$$\begin{aligned} Pr(\mathbf{Y}_{nob}|\mathbf{X}, \mathbf{Y}_{obs}, \mathbf{R}_{inc}, \mathbf{I}) &= Pr(\mathbf{Y}_{nob}|\mathbf{X}, \mathbf{Y}_{obs}) \\ &= Pr(\mathbf{X}, \mathbf{Y}) / \int Pr(\mathbf{X}, \mathbf{Y})d\mathbf{Y}_{nob} \\ &= \int_{\theta} f(\mathbf{X}, \mathbf{Y}|\theta)Pr(\theta)d\theta / \int_{\mathbf{Y}_{nob}} \int_{\theta} f(\mathbf{X}, \mathbf{Y}|\theta)Pr(\theta)d\theta d\mathbf{Y}_{nob} \end{aligned} \quad (15)$$

kde specifikované konkrétní rozdělení $Pr(\mathbf{X}, \mathbf{Y}) = \int f(\mathbf{X}, \mathbf{Y})Pr(\boldsymbol{\theta})d\boldsymbol{\theta}$ je využíváno pro realizaci statistické indukce o sledované proměnné \mathbf{Y} .

V praxi převažuje mechanismus *missing-at-random* (MAR), tj. chybějící data chybí náhodně nezávisle na nepozorovaných (chybějících) složkách proměnné \mathbf{Y} V případě, že nepozorovaná data chybí nenáhodně, tj. mechanismus chybění není náhodný, vztahy (13 až 15) neplatí a je nutné použít jeden z následujících postupů:

- **První přístup představují tzv. modely výběru (selection models):** V prvním kroku specifikuje rozdělení potenciálně kompletních dat (tzn. datové matice \mathbf{Y} , která je složená z pozorovaných a chybějících hodnot). V dalším kroku se podle [13] specifikuje model [5], podle kterého závisí výskyt chybějících hodnot na datové matici \mathbf{Y} .
- **Druhý přístup tvoří tzv. modely smíšených vzorů (pattern-mixture models):** Jednotky jsou rozdělené do skupin podle jedinečných vzorů chybějících hodnot. Následně se v každé ze skupin provádí statistická analýza [4]. Použitý pojem „smíšené vzory“ má indikovat, že výsledné marginální rozdělení datové matice \mathbf{Y} je směsí několika rozdělení [5].

Podrobnější analýza nenáhodného mechanismu chybění v neúplných datech přesahuje rozsah tohoto článku. Detailnější informace o mechanismu NMAR lze dohledat například v [7] nebo v [1].

6. NÁZORNÉ PŘÍKLADY BAYESOVSKÉ ANALÝZY V SYSTÉMU SAS

Programový systém SAS nabízí 2 způsoby analýzy dat bayesovskou metodou. Ve vybraných procedurách pro statistické modelování se bayesovská analýza volá příkazem BAYES a univerzální proceduru MCMC pro obecné statistické modelování pomocí Bayesovy věty [8]. Statistická inference (viz předcházející kapitoly) – bodové i intervalové odhady - vychází z aposteriorního rozdělení. Bayesovské modelování také nabízí alternativní modelové řešení v analýze chybějících hodnot procedurou MI, ve kterém se s chybějícími hodnotami zachází jako s neznámými parametry a podle toho se odhadují [10]. Tyto chybějící hodnoty jsou jednoduše imputovány přijatelnými hodnotami prostřednictvím simulací aposteriorního rozdělení neúplných údajů.

Jádro procedur pro generování pseudonáhodných čísel z pravděpodobnostních rozdělení tvoří metody založené na simulaci pomocí Markovských řetězců. Podle [14] je Markovský řetězec sekvence náhodných veličin, u nichž rozdělení každého prvku závisí výlučně na hodnotě prvku předchozím. Je-li generované rozdělení konjugované, vzorky jsou generovány přímo z podmíněných aposteriorních rozdělení využitím standardních číselných generátorů². Pro účely simulace metodou Markovských řetězců jsou nejpopulárnějšími algoritmy Gibbsovské vzorkování (Gibbs Sampling) a Metropolis-Hastings algoritmus. Metropolis-Hastings algoritmus simuluje výběr z aposteriorního pravděpodobnostního rozdělení za účelem aproximace požadovaného rozdělení a provádí rozhodování o přijetí nebo o odmítnutí vygenerovaného náhodného vzorku. Pokud je nově generovaný vzorek akceptován, algoritmus generuje vzorek nový, pokud je náhodně generovaný vzorek odmítnut,

² Je-li aposteriorní rozdělení členem stejné rodiny rozdělení jako rozdělení apriorní, pak obě tato rozdělení jsou konjugovaná neboli obecně – jsou stejného typu – pozn. aut.

simulace stávajícího výběru se opakuje. Generování náhodných vzorků končí v okamžiku dosažení konvergence k cílovému aposteriori rozdělení. Po dosažení konvergence se vybere náhodně z několika nasimulovaných řad a vybírá se řada podle spolehlivosti konvergence k cílovému rozdělení. Gibbsovo vzorkování (jako speciální případ předchozího algoritmu) rozděluje simultánní aposteriori rozdělení vícerozměrné náhodné veličiny na podmíněná rozdělení pro každý parametr, ze kterých náhodně generuje řetězce a sleduje konvergenci k požadovanému aposteriori rozdělení.

Uvedený příklad se poprvé objevil v práci [6]. Datový soubor představuje náhodně vybraný vzorek s dvojrozměrným normálním rozdělením s chybějícími daty. Jedná se o uměle vytvořený problém, kterým však později zabývali i mnozí další následovníci, například v [17] a [18].

Tabulka č. 2: Struktura souboru s normálním rozdělením s chybějícími daty

Sledované proměnné		Indikátorové (pomocné) proměnné					
X_1	X_2	I		R		M	
1	1	1	1	1	1	1	1
1	-1	1	1	1	1	1	1
-1	1	1	1	1	1	1	1
-1	-1	1	1	1	1	1	1
2	.	1	1	1	0	1	0
2	.	1	1	1	0	1	0
-2	.	1	1	1	0	1	0
-2	.	1	1	1	0	1	0
.	2	1	1	0	1	0	1
.	2	1	1	0	1	0	1
.	-2	1	1	0	1	0	1
.	-2	1	1	0	1	0	1

Zdroj: [4, s. 27]

Soubor dat (nazvaný BINORM) se po provedeném náhodném výběru skládá z 12 pozorování z dvojrozměrného normálního rozdělení. Strukturu souboru s chybějícími daty zobrazuje tabulka č. 2. U obou komponent tohoto vícerozměrného rozdělení proměnných chybí hodnoty. Dvojrozměrné normální rozdělení má obě komponenty se stejnými nulovými průměry μ , s různými rozptyly σ_1^2 a σ_2^2 a korelačním koeficientem ρ . Tyto předpoklady se promítají do následující kovarianční matice Σ :

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \quad (16)$$

Apriorní rozdělení kovarianční matice Σ je $Pr(\Sigma)$ a odpovídá podle [14] inverznímu Wishartovu rozdělení s determinantem $(\sigma_1\sigma_2\sqrt{1-\rho^2})^{-(p+1)}$, tj.

$$Pr(\Sigma) \propto [\Sigma]^{-\frac{p+1}{2}} = (\sigma_1\sigma_2\sqrt{1-\rho^2})^{-(p+1)}, \quad (17)$$

kde symbol p odpovídá počtu sledovaných proměnných, tedy $p = 2$.

Jelikož se jedná o výběrový soubor o velikosti n náhodně vybraný jednoduchým náhodným výběrem, indikátory zahrnutí pro všechny jednotky jsou rovné 1. Mechanismus výběru je proto dán výrazem

$$Pr(\mathbf{I}|\mathbf{X}, \mathbf{Y}, \mathbf{R}) = Pr(\mathbf{I}) = 1/\binom{N}{n} \text{ pro } \sum_{i=1}^N I_i \text{ a } Pr(\mathbf{I}) = 0 \text{ jinak. (18)}$$

Při analýze kompletních pozorování, kdy se do úvahy berou pouze úplná pozorování, by se bayesovským modelováním vytvořil poměrně jednoduchý aposteriorní odhad: 2 datové dvojice ((1,1) a (-1,-1)) mají korelaci rovnou 1 a 2 dvojice dat ((1,-1) a (-1,1)) mají korelaci -1. V případě, že se budou chybějící hodnoty považovat za zcela náhodně chybějící, lze podle [10] považovat nepozorované hodnoty buď (téměř) dokonale pozitivně nebo dokonale negativně korelující s pozorovanými hodnotami 2 a -2 a že chybějící hodnoty mohou nabývat pouze hodnot, které se od nich příliš neliší. Navíc vyřazením částečně pozorovaných hodnot se ztrácí podstatná část při odhadu rozptylu obou komponent dvojrozměrného normálního rozdělení. Bayesovské modelování chybějících hodnot dvojrozměrného normálního rozdělení vykonávají následující příkazy programového systému SAS:

```
proc mcmc data=BiNorm nmc=20000 seed=17 outpost=postout
  diag=none plots=none monitor=(rho sig1 sig2);
  array x[2] x1 x2;
  array mu[2] (0 0);
  array sigma[2, 2];
  parms rho 0.5 / slice;
  parms sig1 1 sig2 1;
  if (sig1 > 0 and sig2 > 0) then
    lprior = -3 * log(sig1 * sig2 * sqrt(1-rho*rho));
    else lprior = .;
  prior rho sig1 sig2 ~ general(lprior);
  sigma[1,1] = sig1*sig1;
  sigma[1,2] = rho * sig1 * sig2;
  sigma[2,1] = sigma[1,2];
  sigma[2,2] = sig2*sig2;
  model x ~ mvn(mu, sigma);
run;
```

Příkazy PARMs specifikují tři modelované parametry modelu. Volba SLICE v prvním příkazu PARMs vybere vzorkovací algoritmus, kterým se generuje parametr ρ . I když výpočetně náročnější [9], tento typ algoritmu je velmi vhodný pro simulaci dat z nenormálních rozdělení (kam náleží i rozdělení korelačního koeficientu ρ). Příkazy IF-ELSE zajišťují, aby simulované hodnoty parametrů směrodatné odchylky σ_1 a σ_2 byly zahrnovány do aposteriorních odhadů kovarianční matice pouze při své kladné hodnotě. Příkazem LPRIOR= a příkazem PRIOR se do procedury zadávají simultánní apriorní rozdělení parametrů σ_1 , σ_2 a ρ na logaritmické stupnici³. Prvky kovarianční matice Σ jsou konstruovány použitím dalších programových příkazů jako funkcí parametrů modelu (v syntaxi příkazů označené jako *sigma*[1,1] až *sigma*[2,2]). Příkazem MODEL se nastavuje specifikace požadované funkce věrohodnosti nutná pro simulaci a následnou aposteriorní inferenci.

³ Použitím výpočtů na logaritmické stupnici se dosahuje větší stability a rychlejší konvergence v číselném výpočtu.

Po provedené simulaci na základě výše uvedené syntaxe je výstupem činnosti procedury MCMC soubor POSTOUT obsahující 20000 vygenerovaných chybějících hodnot v obou proměnných X_1 a X_2 ve dvojrozměrném normálním rozdělení a vygenerované rozdělení parametrů σ_1 , σ_2 a ρ podle zadání. Analýza chybějících hodnot je ilustrována v části *Missing Data Information Table* tabulky č. 3. Proměnná X_1 obsahuje 4 chybějící hodnoty, proměnná X_2 obsahuje také 4 chybějící hodnoty. Tabulka zaznamenává indexy chybějících hodnot za jednotlivé proměnné. Aposteriorní inference o odhadovaných parametrech σ_1 , σ_2 a ρ je zobrazena v části *Posterior Summaries and Intervals* tabulky č. 3.

Tabulka č. 3: Sumární statistiky a inference výsledného rozdělení

Missing Data Information Table				Posterior Summaries and Intervals					
Variable	Number of Missing Obs	Observation Indices	Sampling Method	Parameter	N	Mean	Standard Deviation	95% HPD Interval	
X1	4	9 10 11 12	Direct	rho	20000	-0.00184	0.5582	-0.8591	0.8519
X2	4	5 6 7 8	Direct	sig1	20000	1.6555	0.4458	0.9828	2.6014
				sig2	20000	1.6306	0.4444	0.9478	2.4860

Zdroj: vlastní zpracování

Aposteriorní inference o chybějících hodnotách proměnných X_1 a X_2 je zobrazena ve výstupních tabulkách č. 4.

Tabulka č. 4: Sumární statistiky a inference chybějících hodnot

Parameter	N	Mean	StdDev	P25	P50	P75	Parameter	Alpha	CredibleLower	CredibleUpper	HPDLower	HPDUpper
X2_5	20000	-0.000219	1.84723	-1.25392	0.003381	1.25390	X2_5	0.05	-3.55760	3.61006	-3.61197	3.54433
X2_6	20000	0.001579	1.83949	-1.23536	-0.003435	1.24945	X2_6	0.05	-3.50325	3.53977	-3.48336	3.54871
X2_7	20000	0.004489	1.83727	-1.23108	0.036230	1.22791	X2_7	0.05	-3.60965	3.57439	-3.62454	3.55160
X2_8	20000	-0.008475	1.84616	-1.25881	0.005462	1.23491	X2_8	0.05	-3.58898	3.55509	-3.57001	3.56999
X1_9	20000	-0.023454	1.88411	-1.31367	-0.025875	1.25902	X1_9	0.05	-3.69331	3.62257	-3.75078	3.55437
X1_10	20000	-0.003563	1.88864	-1.27118	0.007136	1.29767	X1_10	0.05	-3.66322	3.64678	-3.63647	3.66321
X1_11	20000	0.003396	1.89574	-1.30166	0.004094	1.27681	X1_11	0.05	-3.69566	3.69085	-3.73594	3.63509
X1_12	20000	0.002021	1.86815	-1.26153	0.018829	1.26093	X1_12	0.05	-3.63658	3.65181	-3.70957	3.55111

Zdroj: vlastní zpracování

Podobně jako v předchozím případě jsou apriorní i podmíněné aposteriorní rozdělení chybějících údajů konjugované. Podmíněná aposteriorní rozdělení chybějících hodnot $Pr(X_1|\cdot)$ a $Pr(X_2|\cdot)$ pro jednotlivé proměnné X_1 a X_2 jsou jednorozměrná normální rozdělení a jsou dána vzorci:

$$Pr(X_1|\rho, \sigma_1, \sigma_2, x_2) \sim N\left(\rho \frac{\sigma_1}{\sigma_2} x_2, \sigma_1^2(1 - \rho^2)\right) \quad (19)$$

$$Pr(X_2|\rho, \sigma_1, \sigma_2, x_1) \sim N\left(\rho \frac{\sigma_2}{\sigma_1} x_1, \sigma_2^2(1 - \rho^2)\right). \quad (20)$$

Následující příklad ilustruje možnosti procedur z programového systému SAS, které jsou přímo určeny k nahrazování chybějících hodnot způsobem zajišťujícím adekvátní statistickou inferenci. Jedná se o procedury MI [15] a MIANALYZE [16], které byly zavedeny jako standardní součásti programového modulu SAS/STAT počínaje verzí 9.0. Vícenásobná imputace se nesnaží odhadnout každou chybějící hodnotu pomocí simulovaných hodnot, ale spíše modeluje náhodný vzorek chybějících hodnot. Výsledek tohoto procesu dává možnost platných statistických závěrů, které řádně

odrážejí nejistotu způsobenou chybějícími hodnotami; například platné intervaly spolehlivosti pro parametry.

Mnohonásobná imputační inference zahrnuje tři různé fáze:

- chybějící údaje se doplní m krát, aby se vytvořilo m úplných souborů dat,
- m úplných datových souborů se analyzuje pomocí standardních postupů a
- výsledky z m úplných datových souborů se zkombinují pro účely inference.

Procedura mnohonásobné imputace v softwaru SAS/STAT je procedura MI a vytváří soubory mnohonásobně imputovaných dat pro neúplná p -rozměrná vícerozměrná data. Procedura používá metody, které zahrnují vhodnou variabilitu napříč m imputacemi. Jakmile je m úplných datových souborů analyzováno pomocí standardních postupů, lze použít proceduru MIANALYZE k vytvoření platných statistických závěrů o těchto parametrech kombinací výsledků z m úplných datových souborů. Tento postup syntetizuje výsledky tím, že vytvoří průměry bodových odhadů, které jsou předmětem zájmu (průměry, odhady parametrů atd.), napříč imputovanými datovými soubory spolu s upravenými rozptyly a směrodatnými chybami, které zohledňují nejistotu vnesenou do procesu imputace.

Výsledný bodový odhad populačního průměru je kombinovaný bodový odhad populačního průměru, což je průměrná hodnota bodových odhadů přes všech m nezávislých opakování imputací, tj.

$$\bar{\theta} = \frac{1}{m} \sum_{i=1}^m \hat{\theta}_i, \quad (21)$$

kde $\hat{\theta}_i$ je odhad parametru θ z i -tého souboru dat (po imputaci) pro $i = 1, \dots, m$.

Vnitroimputační rozptyl (v rámci imputací) je počítán jako průměr odhadovaných rozptylů $var(\hat{\theta}_i)$ bodových odhadů $\hat{\theta}_i$ populačního průměru přes všech m imputací, tj.

$$\bar{W} = \frac{1}{m} \sum_{i=1}^m \hat{W}_i = \frac{1}{m} \sum_{i=1}^m var(\hat{\theta}_i), \quad (22)$$

kde $var(\hat{\theta}_i)$ je odhad rozptylu odhadu populačního průměru pro i -tou imputaci.

Mezimputační rozptyl (mezi imputacemi) je odhadován pomocí vzorce

$$\bar{B} = \frac{1}{m} \sum_{i=1}^m (\hat{\theta}_i - \bar{\theta})^2, \quad (23)$$

Celkový rozptyl odhadu parametru θ po mnohonásobné imputaci je odhadnut pomocí Rubinovy kombinační formule [12], tj.

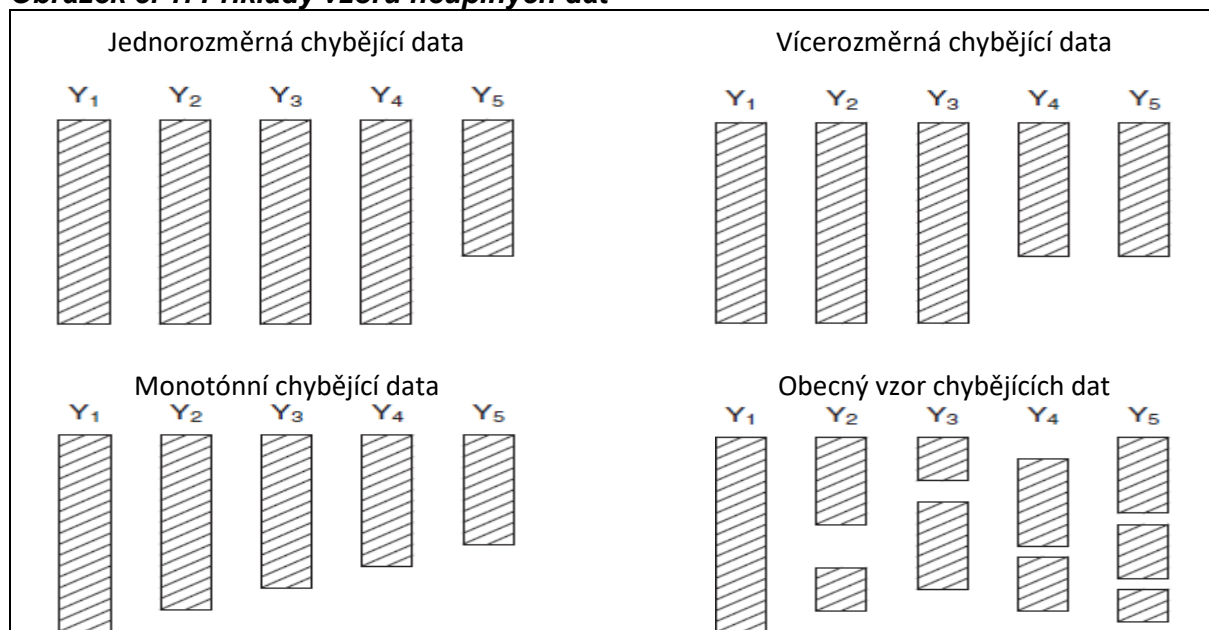
$$var(\bar{\theta}) = \bar{W} + \frac{m+1}{m} \bar{B}. \quad (24)$$

Klíčem ke zvolení imputační metody je statistický model, který definuje pravděpodobnostní vztahy všech uvažovaných proměnných a typ imputované proměnné. Vzor chybějících údajů popisuje umístění chybějících hodnot v datové matici. Za tímto účelem lze řádky a sloupce datové matice seřadit tak, aby vznikly

konkrétní vzory chybějících údajů. Zobrazení 4 speciálních vzorů je ilustrováno na obrázku č. 1.

Například monotónní vzor chybějících údajů znamená, že proměnné datové matice se dají uspořádat tak, aby pro každou další proměnnou chyběly hodnoty u všech jednotek, a mohou chybět i nějaké hodnoty navíc. Tento vzor je automaticky splněn, pokud chybějící hodnoty byly sledovány u jediné proměnné.

Obrázek č. 1: Příklady vzorů neúplných dat



Zdroj: [2, str. 132]

Tabulka č. 5: Shrnutí možných metod imputace pomocí procedury MI podle [15]

Vzor chybění	Typ imputované proměnné	Doporučená metoda imputace
Monotónní	Spojité	Monotónní regrese Monotónní predikovaná shoda v průměru Monotónní propensitní ¹ skóre
	Klasifikační (ordinální)	Monotónní logistická regrese
	Klasifikační (nominální)	Monotónní diskriminační funkce
Libovolný (obecný)	Spojité	Imputace obecným MCMC Imputace monotónní MCMC

¹ Odhad pravděpodobnosti přiřazené každému pozorování u proměnné obsahující chybějící hodnoty, že pozorování je chybějící

Zdroj: vlastní zpracování podle [15]

Shrnutí metod mnohonásobné imputace procedurou MI je uvedeno v tabulce č. 5. Podrobnější informace ke každé z výše uvedených imputačních metod jsou uvedeny především v originální dokumentaci programového systému SAS [15]. Jako příklad využití procedury MI při mnohonásobné imputaci bude prakticky předvedena imputace metodou monotónní regrese pro spojité proměnné s chybějícími pozorováními za podmínek ignorovatelného mechanismu chybění MAR. Příklady mnohonásobných imputací, ve kterých jsou aplikovány ostatní imputační metody, jsou uvedeny v originální dokumentaci pro proceduru MI.

Při regresní metodě se pro každou proměnnou s chybějícími hodnotami sestaví regresní model. Na základě tohoto modelu pro proměnnou Y_j se dále sestaví nový regresní model pro proměnnou Y_{j+1} , který se použije k imputování chybějících hodnot proměnné [12]. Vzhledem k tomu, že soubor dat má monotónní vzor chybějících dat, proces se opakuje postupně pro ostatní proměnné s chybějícími hodnotami.

Pro spojitou proměnnou Y_j s chybějícími hodnotami platí následující regresní model

$$Y_j = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k, \quad (25)$$

který je vyrovnáván na základě pozorovaných hodnot proměnné Y_j a jejích kovariát X_1, X_2, \dots, X_k .

Na základě předchozího textu podle [12] pro proměnnou Y_j s chybějícími hodnotami jsou chybějící hodnoty imputovány z rozdělení

$$Y_j \sim Pr(Y_{nob} | \mathbf{X}, Y_{obs}) = Pr(Y_j | \mathbf{X}, Y_1, Y_2, \dots, Y_{j-1}). \quad (26)$$

K imputování chybějících hodnot pro proměnnou Y_j jsou v každé imputaci uplatněny následující kroky:

- Vyrovnaný model zahrnuje odhady regresních parametrů $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)$ a příslušné kovarianční matice $\hat{\sigma}_j^2 \mathbf{V}_j$, kde kovarianční matice odpovídá inverzní matici $\mathbf{X}'\mathbf{X}$ odvozené z absolutního členu a kovariát X_1, X_2, \dots, X_k .
- Nové parametry $\boldsymbol{\beta} = (\beta_{*0}, \beta_{*1}, \beta_{*2}, \dots, \beta_{*k})$ a σ_{*j}^2 jsou vybírány aposterioriálního predikčního rozdělení parametrů [12], což znamená, že parametry jsou simulovány z $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)$, $\hat{\sigma}_j^2$ a \mathbf{V}_j . Rozptyl je generován jako

$$\sigma_{*j}^2 = \hat{\sigma}_j^2 (n_j - k - 1) / g, \quad (27)$$

kde g je $X_{n_j-k-1}^2$ náhodná proměnná a n_j je počet pozorovaných hodnot v proměnné Y_j . Regresní koeficienty jsou simulovány podle rovnice

$$\boldsymbol{\beta}_* = \hat{\boldsymbol{\beta}} + \sigma_{*j} \mathbf{V}'_{hj} \mathbf{Z}, \quad (28)$$

kde \mathbf{V}'_{hj} , $\mathbf{V}_j = \mathbf{V}'_{hj} \mathbf{V}_{hj}$, je horní trojúhelníková matice daná jako Choleskyho dekompozice a \mathbf{Z} je vektor $k + 1$ nezávislých náhodných proměnných.

Chybějící hodnoty jsou nahrazovány $\beta_{*0} + \beta_{*1} x_1 + \beta_{*2} x_2 + \dots + \beta_{*k} x_k + z_i \sigma_{*j}$, kde x_1, x_2, \dots, x_k , jsou hodnoty kovariát a z_i je i -tá simulovaná náhodná odchylka.

Regresní metodě imputace je velmi podobná metoda shody predikčních průměrů. Rozdíl od regresní metody imputace spočívá v tom, že pro každou chybějící hodnotu imputuje pozorovanou hodnotu, která je nejbližší předpovězené hodnotě ze simulovaného regresního modelu [12].

Pro nominální klasifikační proměnnou je na základě vyrovnaného logistického regresního modelu je podle [15] simulován nový logistický regresní model z posteriorního predikčního rozdělení parametrů a je použit k imputování chybějících hodnot postupně pro každou proměnnou. Logistický regresní model nominální klasifikační proměnné o K úrovních pro imputování chybějících hodnot proměnné Y_j je vyrovnáván pomocí pozorovaných hodnot proměnných zařazených do imputačního modelu. Postupy imputace jsou analogické jako u monotónní lineární regrese u spojité proměnné.

Příklad byl převzat ze zdroje [15]⁴ a ilustruje ukázkou mnohonásobné imputace regresní metodou (logistická regrese pro klasifikační proměnnou Druhy (Species) a lineární regrese pro spojité proměnné výška (Height) a šířka (Width)) nad datovým souborem nazývaném Fish, jehož údaje mají monotónní vzor chybění. Datový soubor obsahuje 2 druhy potočních ryb Cejn (Bream) a Štika (Pike), přičemž každá sledovaná ryba obsahuje informace o délce (Length), výšce (Height) a šířce (Width). Některé hodnoty z těchto 3 měření byly nastaveny jako chybějící a výsledný vzor chybění u proměnných délky (Length), výška (Height), Width (šířka) a Druhy (Species) má monotónní charakter. Prvních 10 pozorování tohoto souboru dat ilustruje tabulka č. 6.

Tabulka č. 6: Soubor dat pro imputaci regresní metodou (prvních 10 pozorování)

Obs	Species	Length	Height	Width
1	Bream	30.0	11.520	4.020
2		31.2	12.480	4.306
3	Bream	31.1	12.378	4.696
4	Bream	33.5	12.730	4.456
5		34.0	12.444	.
6	Bream	34.7	13.602	4.927
7	Bream	34.5	14.180	5.279
8	Bream	35.0	12.670	4.690
9	Bream	35.1	14.005	4.844
10	Bream	36.2	14.227	4.959

Zdroj: vlastní zpracování

Mnohonásobná imputace začíná tvorbou m imputovaných souborů dat pomocí procedury MI. Výsledný soubor dat (kombinace m imputovaných souborů) se nazývá *OutFish*. Sekvence příkazů, které vyvolávají proceduru MI a požadují regresní metodu imputace pro proměnné výška (Height) a šířka (Width) a metodu logistické regrese pro proměnnou Druhy (Species), je následující:

```
proc mi data=Fish seed=1305417 out=OutFish;
  class Species;
  monotone reg(Height Width/ details)
    logistic(Species= Length Height Width Height*Width/
      details);
  var Length Height Width Species;
run;
```

V lineárním regresním modelu bude jako závislá proměnná vystupovat proměnná výška (Height), v modelu logistické regrese je na místě závislé nominální proměnná Druhy (Species). Seznam proměnných zahrnutých do modelu je určen výčtem

⁴ https://documentation.sas.com/doc/en/statcdc/14.3/statug/statug_mi_examples04.htm

proměnných ze souboru dat v příkazu VAR. Pokud příkaz VAR se seznamem proměnných uveden není, do modelu pro imputaci jsou zahrnuty všechny proměnné. Klasifikační proměnná, pokud má být součástí imputačního modelu, musí být uvedena v příkazu CLASS procedury. Po vykonání výše uvedené příkazové sekvence se na výstupu procedury MI objeví výstupní tabulky a další informace uvedené v následujícím textu.

Metodu a možnosti použité v procesu vícenásobné imputace popisuje tabulka *Model Information*. Ve výchozím nastavení se pro chybějící data se vytvoří 25 imputací (počet imputací procedury MI lze změnit příkazem NIMPUTE=). Tabulka *Monotone Model Specification* zobrazuje monotónní metody použité při imputaci (viz tabulka č. 7).

Tabulka č. 7: Informace o imputačním modelu a použitých metodách imputace

Model Information		Monotone Model Specification	
Data Set	WORK.FISH	Method	Imputed Variables
Method	Monotone	Regression	Height Width
Number of Imputations	25	Logistic Regression	Species
Seed for random number generator	1305417		

Zdroj: vlastní zpracování

Jednotlivé vzory chybějících dat znázorňuje tabulka *Missing Data Patterns* v tabulce č. 8 (včetně odpovídacích absolutních a relativních četností). Přítomnost pozorovaných hodnot proměnné v příslušné skupině je indikována znakem „X“. Naopak nevyplněné hodnoty proměnné ilustruje znak „.“. Pro spojitě proměnné je v každé zobrazované skupině zobrazen také průměr proměnné.

Tabulka č. 8: Vzor neúplných dat generovaný procedurou MI

Missing Data Patterns									
Group	Length	Height	Width	Species	Freq	Percent	Group Means		
							Length	Height	Width
1	X	X	X	X	43	82.69	41.997674	12.819512	5.359860
2	X	X	X	.	3	5.77	38.433333	11.797667	4.587667
3	X	X	.	.	4	7.69	42.275000	13.346750	.
4	X	.	.	.	2	3.85	40.150000	.	.

Zdroj: vlastní zpracování

Zobrazit regresní koeficienty v modelu REG, které jsou odhadovány z pozorovaných dat, a regresní koeficienty, které jsou použity v každé imputaci, umožňuje připojení volitelného parametru DETAILS do skupiny příkazů. Ukázka regresních koeficientů simulovaných z aposteriorního rozdělení je v tabulce č. 9 pro prvních 7 iterací.

Tabulka č. 9: Odhady parametrů lineárního modelu pro prvních 7 iterací

Imputed Variable	Effect	Obs-Data							
			1	2	3	4	5	6	7
Height	Intercept	0.00173	-0.152270	-0.136544	-0.064801	0.036585	0.088415	0.125572	0.019478
Height	Length	-0.22453	-0.133455	-0.155687	-0.319043	-0.108935	-0.215399	-0.080855	0.100551

Zdroj: vlastní zpracování

Použitím stejného volitelného parametru DETAILS u logistického imputačního modelu (model LOGISTIC) lze zobrazit hodnoty regresní koeficienty, které jsou odhadovány z pozorovaných dat, a regresní koeficienty, které jsou nasimulovány v každé imputaci. Ukázka koeficientů logistického imputačního modelu je ilustrována tabulkou č. 10.

Tabulka č. 10: Odhady parametrů logistického modelu pro prvních 7 imputací

Imputed Variable	Effect	Obs-Data								
			1	2	3	4	5	6	7	
Species	Intercept	22.80713	22.807129	22.807129	22.807129	22.807129	22.807129	22.807129	22.807129	22.807129
Species	Length	-14.44698	-14.446980	-14.446980	-14.446980	-14.446980	-14.446980	-14.446980	-14.446980	-14.446980
Species	Height	43.11236	43.112363	43.112363	43.112363	43.112363	43.112363	43.112363	43.112363	43.112363
Species	Width	-9.64352	-9.643524	-9.643524	-9.643524	-9.643524	-9.643524	-9.643524	-9.643524	-9.643524
Species	Height*Width	-9.73015	-9.730154	-9.730154	-9.730154	-9.730154	-9.730154	-9.730154	-9.730154	-9.730154

Zdroj: vlastní zpracování

Výstupem procedury MI je soubor dat nazvaný *OutFish*, jenž obsahuje kombinaci m imputovaných souborů. Prvních 10 pozorování tohoto souboru dat ilustruje tabulka č. 11. Z obrázku je patrné, že procedura MI nahradila chybějící hodnoty imputovanými, například do druhého pozorování byla doplněna hodnota proměnné Species, a to již v rámci prvního imputovaného souboru.

Tabulka č. 11: Soubor dat OutFish na výstupu procedury MI (prvních 10 pozorování)

Obs	Imputation	Species	Length	Height	Width
1		1 Bream	30.0	11.520	4.02000
2		1 Bream	31.2	12.480	4.30600
3		1 Bream	31.1	12.378	4.69600
4		1 Bream	33.5	12.730	4.45600
5		1 Bream	34.0	12.444	4.62964
6		1 Bream	34.7	13.602	4.92700
7		1 Bream	34.5	14.180	5.27900
8		1 Bream	35.0	12.670	4.69000
9		1 Bream	35.1	14.005	4.84400
10		1 Bream	36.2	14.227	4.95900

Zdroj: vlastní zpracování

Výstupem procedury MI je také informace o rozptylu, a to jak mezi imputovanými soubory, tak i v rámci imputovaných souborů, který je potřebný k následné inferenci procedurou MIANALYZE. Informace o rozptylech je zaznamenána v tabulce č. 12.

Tabulka č. 12: Odhad rozptylu v rámci imputovaných souborů a mezi soubory imputace

Variance Information (25 Imputations)							
Variable	Variance			DF	Relative Increase in Variance	Fraction Missing Information	Relative Efficiency
	Between	Within	Total				
Height	0.010482	0.296800	0.307701	47.254	0.036730	0.035529	0.998581
Width	0.000534	0.014490	0.015045	47.171	0.038339	0.037032	0.998521

Zdroj: vlastní zpracování

Po vytvoření m imputovaných souborů pomocí procedury MI se přechází lze přejít ke standardní statistické analýze uvedených m imputovaných souborů. Následující

příkazová sekvence vytváří datové soubory, které obsahují odhady parametrů a odpovídající kovarianční matice odhadovaných pomocí logistické regrese. Pomocí standardní procedury LOGISTIC se pro každý imputovaný soubor generují odhady parametrů logistického imputačního modelu (v proceduře MI je tento model zadaný příkazem LOGISTIC). Všechny m imputovaných souborů je uloženo v souboru *OutFish* z výstupu procedury MI:

```
proc logistic data=OutFish;
  class Species;
  model Species= Length Height Width Height*Width / covb;
  by _Imputation_;
  ods output ParameterEstimates=lgpparms CovB=lgcovb;
run;
```

Výstupem procedury LOGISTIC je soubor s odhady parametrů a odhady jejich standardních chyb nazvaný LGPARMS pro všech m imputovaných souborů. Odhady kovariančních matic všech m imputačních modelů jsou uloženy ve výstupním souboru LGCOVB. Oba tyto soubory odhadů parametrů a kovariančních matic z výstupu procedury LOGISTIC jsou vstupními soubory do procedury MIANALYZE v poslední fázi procesu imputačního procesu - kombinace výsledků imputace m imputovaných souborů a celkové statistické inference. Podobným způsobem lze analyzovat i lineární regresní model z procedury MI (v proceduře MI je tento model zadaný příkazem REG).

Závěrečnou fází imputačního procesu je použití procedury MIANALYZE, jejíž pomocí lze výsledky z m imputovaných (úplných) datových souborů se zkombinovat pro účely inference. K tomuto účelu lze použít následující sekvenci příkazů:

```
proc mianalyze parms=lgpparms covb=lgcovb;
  modeleffects Intercept Length Height Width Height*Width;
run;
```

Vstupními soubory pro proceduru MIANALYZE jsou soubory s odhady parametrů (LGPARMS) a odhadů kovariančních matic (LGCOVB) pro všech m imputovaných (úplných) datových souborů z výstupu analytické procedury LOGISTIC, tj. soubory se vloží do procedury MIANALYZE pomocí příkazů PARMS= a COVB=. Příkaz MODELEFFECTS umožňuje analyzovat vliv jednotlivých efektů (nezávislých proměnných) v imputačním modelu.

Výstup procedury MIANALYZE představují informace, které slouží jako výsledná statistická inference po realizovaném procesu imputace. Procedura MIANALYZE kombinuje vstupní informace všech ze všech m imputovaných souborů do výsledných odhadů a zajišťuje tak statistickou inferenci celého procesu imputace. Tabulka *Model Information* obsahuje informace o datových souborech vstupujících do analýzy procedurou MIANALYZE a údaje o počtu imputací. Odhady celkového rozptylu včetně složek rozptylu jsou zaznamenány v tabulce *Variance Information*. Přehled odhadů parametrů imputačního modelu včetně odhadů standardních chyb regresních koeficientů obsahuje tabulka *Parameter Estimates*. Ukázkou statistické inference procedurou MIANALYZE ilustruje tabulka č. 13.

Tabulka č. 13: Ukázka výstupů procedury MIANALYZE po procesu imputace

Model Information								
PARMS Data Set			WORK.LGPARMS					
COVB Data Set			WORK.LGCOVB					
Number of Imputations			25					

Variance Information (25 Imputations)									
Parameter	Variance			DF	Relative Increase in Variance	Fraction Missing Information	Relative Efficiency		
	Between	Within	Total						
Intercept	400.704089	180035	180452	4.5E6	0.002315	0.002310	0.999908		
Length	0.080968	38.419197	38.503404	5.02E6	0.002192	0.002187	0.999913		
Height	2.408527	1136.358955	1138.863824	4.96E6	0.002204	0.002200	0.999912		
Width	32.413980	9735.783085	9769.493624	2.02E6	0.003463	0.003452	0.999862		
Height*Width	0.077975	33.699632	33.780726	4.16E6	0.002406	0.002401	0.999904		

Parameter Estimates (25 Imputations)										
Parameter	Estimate	Std Error	95% Confidence Limits		DF	Minimum	Maximum	Theta0	t for H0: Parameter=Theta0 Pr > t	
Intercept	-39.590964	424.796213	-872.176	792.9945	4.5E6	-125.766370	-22.874243	0	-0.09	0.9257
Length	-0.542283	6.205111	-12.704	11.6195	5.02E6	-1.379808	-0.139264	0	-0.09	0.9304
Height	5.688145	33.747057	-60.455	71.8312	4.96E6	4.215492	12.157530	0	0.17	0.8661
Width	7.243556	98.840749	-186.481	200.9680	2.02E6	-0.014259	24.732078	0	0.07	0.9416
Height*Width	-0.650649	5.812119	-12.042	10.7409	4.16E6	-1.767419	-0.392503	0	-0.11	0.9109

Zdroj: vlastní zpracování

7. ZÁVĚR

V současné době jsou neúplné (chybějící) údaje velmi častou součástí souborů dat, přičemž jejich výskytu není přiřkládán dostatečný význam. Většina výzkumníků při obvyklých analýzách dat zaznamenaných ve statistickém šetření vypouští zpravodajské jednotky s chybějícími údaji z analýz bez ohledu na mechanismus, který chybějící data generuje. Nepříznivé důsledky takového postupu v odhadování parametrů populace je možné do určité míry kompenzovat využitím informací o populaci (pokud jsou k dispozici).

Při řešení problémů imputace chybějících dat by výzkumníci měli zvažovat moderní statistické přístupy vycházející z bayesovského modelování chybějících dat. I když bayesovské metody ve srovnání s klasickými postupy vyžadují hlubší znalosti teorie současně s určitými výpočetními dovednostmi, jejich správná aplikace ve výzkumné činnosti se rozhodně vyplatí. Bayesovská statistika poskytuje obecnější přístup pro statistickou inferenci. Na rozdíl od četnostní statistiky bayesovský přístup považuje odhadované parametry jako náhodné, jejichž rozdělení je možné upřesňovat informacemi získanými z dat. Navíc v kontextu bayesovského přístupu jsou znalosti zkoumaného fenoménu, které lze efektivně využít, zpravidla vždy dostupné. V současné době je aplikace bayesovských metod mnohem přístupnější, protože už existují vhodné volně přístupné programy i placené programové systémy, například právě SAS.

Zvolené příklady do tohoto příspěvku patří k těm jednodušším imputačním problémům, které se však ve statistické praxi vyskytují nejčastěji. Je jasné, že složitější typy imputací vyžadují sofistikovanější statistické přístupy. Cílem zvolených příkladů

bylo přiblížit podstatu nahrazování chybějících hodnot s důrazem na úlohu bayesovského modelování. Dalším cílem článku je ilustrovat možnosti programového systému SAS při řešení náhrady imputacemi chybějících dat. Bližší popis a možnosti procedur SAS pro aplikaci imputačních metod bychom chtěli uveřejňovat v následujících vydáních, neboť problematika imputací a jejich řešení v programovém systému SAS je tak rozsáhlá, že není možné vše obsáhnout jediným příspěvkem.

LITERATURA

- [1] ALLISON, P. D.: Missing Data. Thousand Oaks, CA: Sage. Sage University Papers Series. Quantitative Applications in the Social Sciences. No. 07-136. 2001. ISBN 0-7619-1672-5.
- [2] BOX, G., E. P. – TIAO, G. C.: Bayesian Inference in Statistical Analysis. 1973. New York: John Wiley & Sons.
- [3] GAVALAKIS L. – KONTOYIANNIS I.: Information-theoretic de Finetti-style theorems. 2022 IEEE Information Theory Workshop (ITW).
- [4] ENDERS, C., K.: Applied Missing Data Analysis, Second Edition. New York: Guilford Press, 2022. ISBN 978-1-462-54986-3.
- [5] GRAHAM, J., W.: Missing data: Analysis and design. 2010. New York: Springer. ISBN 978-1-4614-4017-8.
- [6] MURRAY, G. D.: Discussion of Paper by Dempster, Laird, and Rubin. In: Journal of the Royal Statistical Society, Series B, 1977. 39, pp. 27 – 28.
- [7] LITTLE, R., J.A.: Pattern-Mixture Models for Multivariate Incomplete Data. In: Journal of the American Statistical Association, 1993. No. 421. pp. 125 – 134.
- [8] NAKAGAWA, Sh.: Missing data: Mechanisms, methods, and messages. In G. A. Fox, S. Negrete-Yankelevich, & V. J. Sosa (Eds.), Ecological statistics: contemporary theory and application. 2015. pp. 81 – 105.
- [9] NEAL, R. M.: Slice Sampling. 2000. In: Technical Report No. 2005. Department of Statistics and Department of Computer Science. University of Toronto, Toronto, Ontario, Canada. Dostupné na: <https://www.cs.toronto.edu/~radford/ftp/slc-samp.pdf>.
- [10] CHEN, F.: Practical Bayesian Computation using SAS. In: ASA Conference on Statistical Practices. 2014. February 20, s.3.
- [11] PAVELKA, R.: Statistická analýza chybějících dat. In: Slovenská štatistika a demografia. 2024. č. 2. pp. 3 – 25.
- [12] RUBIN, D. B.: Multiple imputation for nonresponse in surveys. New York: Springer. 1987. ISBN 0-471-08705-X.
- [13] RUBIN, D. B.: Inference and Missing Data. In: Biometrika, 1976. Vol. 63, No. 3 pp. 581 – 592.
- [14] SAS Institute Inc.: SAS/STAT® 15.3 User's Guide. 2023. Chapter 8 Introduction to Bayesian Analysis Procedures. Cary, NC: SAS Institute Inc.
- [15] SAS Institute Inc. SAS/STAT® 15.3 User's Guide. 2023. Chapter 82 The MI Procedure. Cary, NC: SAS Institute Inc.
- [16] SAS Institute Inc.: SAS/STAT® 15.3 User's Guide. 2023. Chapter 83 The MIANALYZE Procedure, Cary, NC: SAS Institute Inc.
- [17] TAN, M., T. – TIAN, G-L. – NG, K., W., eds.: Bayesian Missing Data Problems: EM, Data Augmentation, and Non-iterative Computation, 2010. New York: Chapman & Hall/CRC.
- [18] TANNER, M. A. – WONG, W., H.: The Calculation of Posterior Distributions by Data Augmentation. In: Journal of the American Statistical Association, 82, pp. 528 – 540.

RESUMÉ

Zejména pro analýzy metodami vícerozměrných statistik představují chybějící údaje problém. Ačkoliv neúplná data ve výběrovém souboru mohou být zastoupena v relativně malém procentu, může tato situace v zjištěných datech vyústit v relativně velmi malý soubor s kompletními údaji; zejména v případě, kdy u různých jednotek chybí hodnoty různých veličin. V běžné praxi výběrových zjišťování jednotky, u kterých byly zaznamenány nevyplněné hodnoty zjišťovaných ukazatelů, jsou převážně z dalších analýz vyloučeny. Vynechání jednotek z analýz může mít značné negativní dopady – snížení přesnosti odhadů a síly vykonávaných statistických testů a může vést až ke zkresleným výsledkům nevhodných k zobecňování na cílovou populaci.

Bayesovské modelování – náhodné výběry z aposterioriho rozdělení, simulace pomocí metody Monte Carlo a Markovských řetězců-představují pevný základ téměř celé kategorie parametrických metod imputace. I když nejdůležitější faktor v podmínkách neúplných dat představuje mechanismus chybějících hodnot, který má nejvýraznější vliv na úspěšnost rozmanitých metod práce s chybějícími hodnotami, znalost a správné použití metod statistického modelování vycházející z daného konkrétního výběru (vzorku) v kombinaci s apriorními znalostmi či informacemi je významným předpokladem pro korektní nahrazování chybějících údajů imputovanými daty.

RESUME

Missing data is especially a problem for analyses using multivariate statistical methods. Although the incomplete data in the random sample may be represented by a relatively small percentage, this situation may result in a relatively very small set of complete data in the observed data; especially when the values of different quantities are missing for different units. In the normal practice of sample surveys of these units, for which unfilled values of the surveyed indicators were recorded, they are mostly excluded from further analyses. Removing units from analyses can have significant negative effects - reducing the accuracy of estimates and the power of performed statistical tests - and can lead to distorted results unsuitable for generalization to the target population.

Bayesian modeling-random sampling from the posterior distribution, Monte Carlo simulation, and Markov chains-forms a solid foundation for almost the entire category of parametric imputation methods. Although the most important factor under incomplete data conditions is the mechanism of missing value, which has the most significant impact on the success of the various methods for working with the missing values, knowledge and correct application of statistical modelling methods based on a given specific sample, combined with a priori knowledge or information, is an important prerequisite for a correct replacement of missing data with the imputed data.

PROFESIJNÝ ŽIVOTOPIS

Ing. Roman Pavelka, PhD., v letech 1995 – 2010 pracoval v poradenské společnosti Trexima, s. r. o. Na pozici statistik – analytik se zabýval analýzami zejména mzdových a personálních dat. Podílel se na tvorbě pravidelných statistických přehledů a reportů. Spolupracoval s akademickými pracovišti, agenturami i soukromými subjekty na realizaci a vyhodnocování ad hoc statistických výzkumů. Oblast jeho vědeckého zájmu představují výběrová šetření, odhady a statistické modely. V letech 2012 až 2013 se zúčastnil zahraniční stáže ve Velké Británii. Od roku 2013 působil v Národnom ústave certifikovaných meraní vzdelávania (NÚCEM), kde zajišťoval statistické vyhodnocování výsledků testování žáků a studentů. Od roku 2015 pracuje v odboru metod statistických zjišťování Štatistického úradu SR.

KONTAKT

roman.pavelka@statistics.sk

Informatívny článok/Informative article

ŽENY NA TRHU PRÁCE NA SLOVENSKU A VO VEREJNOM ŽIVOTE V POROVNANÍ S KRAJINAMI EURÓPSKEJ ÚNIE

WOMEN IN THE LABOUR MARKET IN SLOVAKIA AND IN PUBLIC LIFE COMPARED TO THE COUNTRIES OF THE EUROPEAN UNION

ABSTRAKT

Problematikou postavenia slovenských žien sa Štatistický úrad Slovenskej republiky zaoberá vo viacerých oblastiach. Cieľom tohto článku je opísať a analyzovať faktory ovplyvňujúce postavenie žien v spoločnosti na Slovensku, predovšetkým na trhu práce, v oblasti odmeňovania za prácu a vo verejnom živote a porovnať ich s vývojom v krajinách Európskej únie. Článok sa zaoberá aj mierou úspešnosti žien vo voľbách do Európskeho parlamentu, národného parlamentu a do orgánov samosprávy obcí, miest a krajov v SR.

ABSTRACT

The Statistical Office of the Slovak Republic deals with the issue of the Slovak women's position in several fields. The aim of this article is to describe and analyse the factors influencing the position of women in the society in Slovakia, mainly in the labour market, in the field of remuneration for work and in the public life and to compare their situation to the development in the EU countries. The article deals too with the success rate of women in the elections to the European Parliament, the national parliament and the self-governing bodies of municipalities, cities and regions in Slovakia.

KLÚČOVÉ SLOVÁ

rovnosť medzi ženami a mužmi, miera ekonomickej aktivity, miera zamestnanosti, miera nezamestnanosti, miera rizika príjmovej chudoby, rodový mzdový rozdiel, verejný život

KEY WORDS

equality between women and men, economic activity rate, employment rate, unemployment rate, at risk of poverty rate, gender pay gap, public life

1. ÚVOD

Rovnosť ľudí je ideálom, ku ktorému sa v súčasnosti hlásia všetky civilizované a demokratické krajiny sveta a spoločenstvá týchto krajín. Európska únia (EÚ) zakotvuje rovnosť pohlaví medzi základné princípy a zaväzuje členské štáty uplatňovať vo všetkých politikách opatrenia na jej posilňovanie.

Štatistický úrad SR dlhodobo zisťuje, spracúva a poskytuje štatistické informácie v členení za ženy a mužov. Tým umožňuje opísať javy ovplyvňujúce postavenie žien a mužov v spoločnosti a napomáha pri tvorbe a prijímaní opatrení na zvýšenie úrovne podpory, ochrany a dodržiavania rovnosti práv a príležitostí medzi pohlaviami.

Pracovisko Štatistického úradu SR v Košiciach od roku 2010 vydáva publikáciu Rodová rovnosť v SR, teraz s aktualizovaným obsahom a názvom **Ženy a muži v SR**, zameranú na porovnanie rozdielov medzi ženami a mužmi v najrôznejších oblastiach života

v modernej spoločnosti. Publikáciu je možné získať na webovom sídle Štatistického úradu SR www.statistics.sk.

V zhruba polovici krajín sveta (asi 55 %) v súčasnosti početne prevládajú ženy nad mužmi, v necelej tretine krajín, najmä v rozvojových, muži nad ženami. V ostatných krajinách je pomer pohlaví pomerne vyrovnaný. V krajinách EÚ27 na 100 mužov pripadá 104,6 žien. Podobne je to aj na Slovensku 104,5 žien (r. 2023) a podiel ženskej zložky obyvateľstva Slovenska tvorí 51,1 %.

Ženy a muži majú mať rovnakú moc formovať spoločnosť a svoj vlastný život, čo znamená:

- rovnomerné rozdelenie moci a vplyvu pri rozhodovaní,
- ekonomickú rovnosť medzi ženami a mužmi,
- rovnomerné rozdelenie neplatennej starostlivosti a práce v domácnosti,
- koniec násillia mužov voči ženám.

Doterajší vývoj na Slovensku v oblasti rovnosti medzi ženami a mužmi ovplyvňuje celý rad faktorov, ku ktorým nepochybne patrí aj tradičné rozdelenie rodových rolí, nízka znalosť svojich práv a agendy rodovej rovnosti, nedostatok služieb na zosúladienie pracovného a rodinného života.

2. ŽENY NA TRHU PRÁCE

Jednou z oblastí, kde sa prejavuje pomerne výrazná nerovnosť medzi ženami a mužmi, je **zamestnanosť a mzdy**.

Zamestnanosť žien na Slovensku je porovnateľná s krajinami EÚ27 (tabuľka č. 1), podiel pracujúcich žien v SR predstavuje 47,1 %, oproti 46,6 % pracujúcich žien v európskych krajinách (vo veku 15 – 64 rokov). Ovpływujú ju však špecifické faktory, ako sú väčšie regionálne rozdiely a málo pracovných miest na kratšie úvázky. **Miera ekonomickej aktivity** (vo veku 15 – 64 r.) u žien v roku 2023 dosiahla 72,8 %, čo je o 7,4 percentuálneho bodu (p. b.) menej ako u mužov, napriek skutočnosti, že počet žien v produktívnom veku (15 – 64 r.) bol podľa výsledkov výberového zisťovania pracovných síl (VZPS) v roku 2023 len o 33,3 tisíca, resp. 1,9 p. b. nižší ako počet mužov v rovnakej vekovej skupine. V krajinách EÚ27 v roku 2023 dosiahol rozdiel mier ekonomickej aktivity medzi ženami a mužmi (vo veku 15 – 64 rokov) 9,6 p. b. Medziročne sa v SR zaznamenal mierny nárast miery ekonomickej aktivity u oboch pohlaví (u mužov o 0,4 p. b., u žien o 0,6 p. b.).

Podobný vývoj sa zaznamenal aj v rozdieloch v **mierach zamestnanosti** medzi mužmi a ženami (vo veku 15 – 64 rokov), ktorý v roku 2023 dosiahol 7,1 p. b. v neprospech žien (medziročne sa tento rozdiel zmiernil o 0,3 p. b.). Miera zamestnanosti v SR sa oproti roku 2022 zvýšila u žien o 0,8 p. b. a u mužov o 0,5 p. b. V krajinách EÚ27 bol rozdiel mier zamestnanosti 9,4 p. b. v neprospech žien, pri medziročnom poklese o 0,5 p. b., pričom nárast miery zamestnanosti sa zaznamenal u oboch pohlaví, výraznejšie u žien (o 0,7 p. b.).

Slovenská republika dlhodobo patrí v Európe k tým krajinám, kde je najnižší podiel osôb zamestnaných na **kratší pracovný čas**. Ženy mávajú skrátený pracovný čas a často pracujú na čiastočný úväzok, aby mohli lepšie skombinovať rodinné povinnosti s platenou prácou. Efekt znevýhodňovania žien v dôsledku nižšieho pracovného úväzku sa na Slovensku neprejavuje tak výrazne ako v iných krajinách. Kým v krajinách EÚ pracuje na kratší pracovný čas približne každý desiaty muž a každá tretia žena, v SR túto formu pracovného úväzku využívalo v roku 2023 iba 2,4 % mužov a 5,4 % žien. Medziročne sa tento podiel zvýšil iba u žien, o 0,8 p. b.

Tabuľka č. 1: Vybrané indikátory rovnosti medzi ženami a mužmi

Indikátor	Ženy		Muži		Rozdiel	
	SR	EÚ27	SR	EÚ27	SR	EÚ27
Trh práce						
Miera zamestnanosti (v %, 20 – 64 rokov, 2023)	73,6	70,2	81,3	80,4	7,7	10,2
Miera zamestnanosti (v %, 15 – 64 rokov, 2023)	68,4	65,7	75,5	75,1	7,1	9,4
Miera zamestnanosti mladých (v %, 15 – 24 rokov, 2023)	17,8	33,0	25,4	37,3	7,6	4,3
Miera zamestnanosti starších (v %, 55 – 64 rokov, 2023)	63,4	58,1	70,1	70,1	6,7	12,0
Miera nezamestnanosti (v %, všetky vekové skupiny, 2023)	5,9	6,4	5,8	5,8	-0,1	-0,6
Miera dlhodobej nezamestnanosti (12 a viac mesiacov, v %, 2023)	3,9	2,2	3,7	2,0	-0,2	-0,2
Miera nezamestnanosti mladých (v %, 15 – 24 rokov, 2023)	16,0	14,0	22,1	14,9	6,1	0,9
Príjem a vzdelanie						
Mzdový rozdiel medzi ženami a mužmi (v % hodinovej hrubej mzdy, 2022)					17,7	12,7
Podiel populácie vo veku 30 – 34 rokov s vysokoškolským vzdelaním (v %, ISCED 5-8, 2022)	51,6	49,4	30,9	38,5	-20,7	-10,9
Skorý odchod zo vzdelávania – populácia vo veku 18 – 24 rokov s nižším ako stredoškolským vzdelaním (v %, ISCED 0-2, 2022)	6,0	7,7	6,8	11,3	0,8	3,6
Mladí ľudia vo veku 15 – 24 rokov mimo zamestnania i vzdelávania – NEET (v %, 2022)	9,1	9,7	10,1	9,5	1,0	-0,2
Chudoba						
Miera rizika príjmovej chudoby (v %, 2023)	13,8	17,0	14,7	15,4	0,9	-1,6
Miera rizika príjmovej chudoby starších (v %, 65 a viac rokov, 2023)	11,0	19,0	7,6	14,0	-3,4	-5,0
Miera rizika chudoby alebo sociálneho vylúčenia jednorodičovských rodín (v %, najmenej 1 dieťa, 2022) *					46,5	43,5
Demografia						
Stredná dĺžka života pri narodení (v rokoch, 2023)	80,5	83,3	73,6	77,9	-6,9	-5,4
Stredná dĺžka života pri narodení v zdraví (v rokoch, SR 2021, Eurostat)	57,5	64,2	56,2	63,1	-1,3	-1,1
Index rovnosti žien a mužov (v %, 2023, EIGE)					59,2	70,2

* ide o údaj za všetky domácnosti daného zloženia (bez členenia podľa pohlavia)

Poznámka: rozdiel medzi ženami a mužmi = hodnoty mužov mínus hodnoty žien v percentuálnych bodoch, ak nie je uvedené inak.

Zdroj: Štatistický úrad SR, Eurostat, EIGE

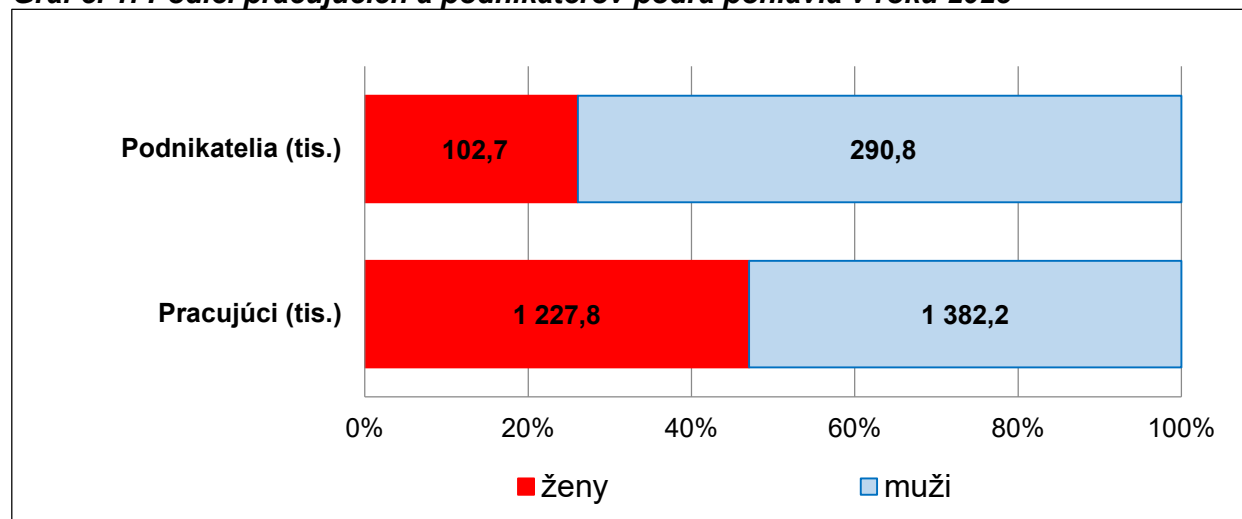
Podľa výsledkov VZPS za rok 2023 v **miere nezamestnanosti** na Slovensku bol rozdiel medzi mužmi a ženami len nepatrný, a to 0,1 p. b. v neprospech žien. U žien sa zaznamenala v priemere 5,9 % nezamestnanosť, čo je o 0,5 p. b. pod európskou

úrovňou, u mužov dosiahla 5,8 %, rovnako ako je priemer za európske krajiny. Dlhodobá nezamestnanosť (12 a viac mesiacov) poklesla od roku 2009 na najnižšiu úroveň, v európskych krajinách na 2,1 % a na Slovensku na 3,8 %, čo je druhá najnižšia miera za posledných 15 rokov. Vyššiu mieru dlhodobej nezamestnanosti v SR zaznamenali ženy, a to 3,9 % (o 0,2 p. b. vyššiu než muži).

V pracovnom procese sú ženy konfrontované s horizontálnou i vertikálnou segregáciou, čo je stav, keď ženy pracujú v odlišných a častejšie v menej platených zamestnaniach ako muži (horizontálna segregácia) a pracujú na nižších pracovných pozíciách (vertikálna segregácia). Vo verejnom sektore ženy predstavujú vyše 62 % zamestnancov, naopak, v súkromnom sektore s vyššími zárobkami ženy predstavujú necelých 45 % všetkých zamestnancov (graf č. 1). V oblastiach ako zdravotníctvo, sociálna pomoc a vzdelávanie (kde sú nižšie priemerné zárobky) ženy tvoria viac než 82 % pracovnej sily, v sektore verejnej správy 60 %.

Rozdiely medzi mužmi a ženami v SR sa najvýraznejšie prejavili v podnikateľskej sfére, kde na jednu ženu podnikateľku pripadajú približne traja muži podnikatelia (graf č. 1). Napriek tomu, že počet žien podnikateliek má stúpajúci trend, rodový rozdiel v SR v roku 2023 dosiahol 7,2 p. b. v neprospech žien. V krajinách EÚ27 na jednu ženu podnikateľku pripadajú dvaja muži podnikatelia a rodový rozdiel predstavuje 4,8 p. b.

Graf č. 1: Podiel pracujúcich a podnikateľov podľa pohlavia v roku 2023



Zdroj: Štatistický úrad SR, výberové zisťovanie pracovných síl, 2023

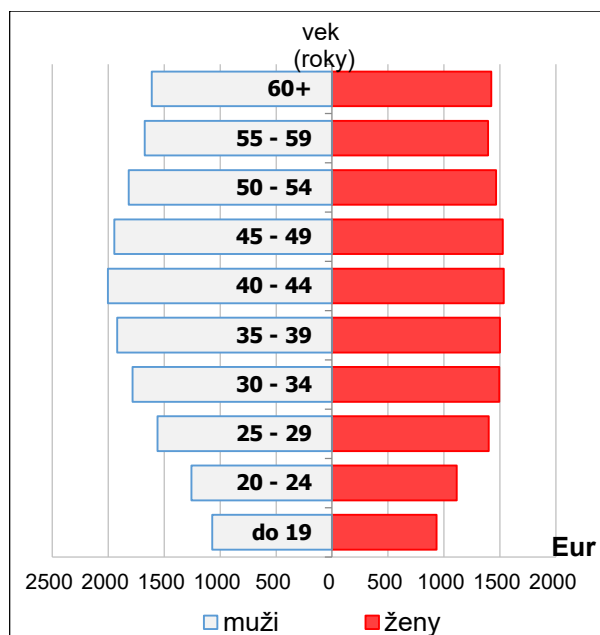
3. MZDOVÝ ROZDIEL MEDZI ŽENAMI A MUŽMI

Rozdiely v odmeňovaní žien a mužov predstavujú komplexný problém zapríčinený viacerými súvisiacimi faktormi. Patria k nim osobné charakteristiky, vek a prax zamestnanca, investície do ľudského kapitálu – vzdelanie, charakteristiky na trhu práce (typ pracovného úväzku, zmluvy), firemné charakteristiky (odvetvie, veľkosť firmy, región, druh vlastníctva) a s nimi súvisiaca horizontálna segregácia (podiel žien vo firme, v odvetví, zamestnaní).

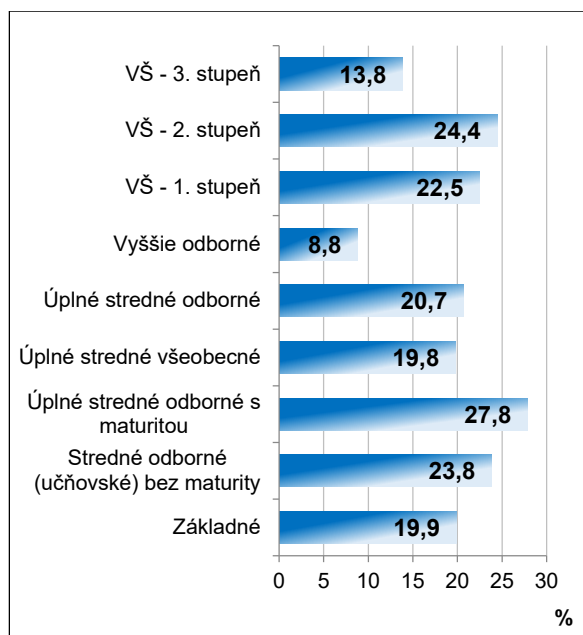
Napriek zaznamenanému pokroku v tejto oblasti stále pretrvávajú výrazné rozdiely v platovom ohodnotení žien a mužov, pričom v EÚ patríme ku krajinám s jeho najvyššou hodnotou. Pozitívnu správou je, že sa mzdový rozdiel medzi pohlaviami na Slovensku postupne znižuje.

Mzdový rozdiel medzi ženami a mužmi (vypočítaný z priemernej hrubej mesačnej mzdy, bez ohľadu na dĺžku pracovného úväzku) od roku 2020 poklesol pod 20 %-nú hranicu a v roku 2023 dosiahol 18,4 %.¹ Ženy v roku 2023 zarábali v priemere 1 462 Eur a ich mzdy stúpili oproti predchádzajúcemu roku o 8,9 %, mzdy mužov medziročne stúpili o 8,3 % na 1 792 Eur. Napriek tomu sa mzda žien nachádzala o 11,8 % pod celoslovenským priemerom, zatiaľ čo mzda mužov ho prevýšila o 10,1 %. Najvýraznejšie sa tieto mzdové rozdiely prejavili vo vekovej kategórii u 35 – 39-ročných, 471 Eur a 40 – 44-ročných, 421 Eur (graf č. 2). Z hľadiska najvyššie dosiahnutého vzdelania boli najvyššie rozdiely v priemernej hrubej mesačnej mzde mužov oproti ženám pri dosiahnutí úplného stredného odborného vzdelania s maturitou (graf č. 3). V krajinách EÚ27 rodový mzdový rozdiel v neočistenej forme² v roku 2022 predstavoval 12,7 % v neprospech žien, čo bolo o 5 p. b. menej než na Slovensku.

Graf č. 2: Priemerná hrubá nominálna mesačná mzda podľa veku a pohlavia v roku 2023



Graf č. 3: Rozdiely priemernej hrubej mesačnej mzdy mužov oproti ženám podľa vzdelania v roku 2023



Zdroj: Štatistický úrad SR, výberové zisťovanie o štruktúre miezd

¹ Z výsledkov výberového zisťovania o cene práce realizovaného spracovateľskou organizáciou Trexima, s r. o. (ISCP 1-04) za rok 2023.

² Podľa definície Eurostatu je to rozdiel medzi priemernou hrubou hodinovou mzdou mužského plateného zamestnanca a ženskou platenou zamestnankyňou ako percento hrubej hodinovej mzdy mužského zamestnanca. Zisťovanie sa uskutočnilo v podnikoch s 10 a viac zamestnancami podľa NACE.

Ženy častejšie pracujú v odvetviach s výrazne nižším finančným ohodnotením, k tomu sa pridáva priama diskriminácia v odmeňovaní, založená na rodovo stereotypných presvedčeniach zamestnávateľov týkajúcich sa napr. dominantnej roly muža – živateľa v rodine. Napriek veľmi dobrej vzdelanostnej úrovni slovenských žien (vyššej než u mužov) ženy v priemere nedosahujú porovnateľné zárobky s mužmi, keďže vzdelanie v tzv. typicky ženských oblastiach je oceňované nižšie.

Rozdiely v odmeňovaní žien a mužov zapríčiňujú, že ženy majú nižšie celoživotné zárobky. V dôsledku toho dostávajú nižšie dôchodky a existuje u nich vyššie riziko chudoby v starobe. V roku 2023 bola podľa zisťovaní EU SILC miera rizika príjmovej chudoby starších žien (vo veku 65 rokov a viac) 11 % a v tej istej vekovej kategórii u mužov 7,6 %, t. j. rozdiel 3,4 p. b. v neprospech žien. V európskych krajinách je tento rodový rozdiel vyšší v porovnaní s vývojom na Slovensku, zaznamenal 5 p. b. v neprospech žien pri 19 % miere rizika chudoby starších žien.

4. ŽENY A MUŽI VO VEREJNOM ŽIVOTE A V ROZHODOVANÍ

Na prehĺbovanie mzdového rozdielu v neprospech žien negatívne vplýva aj ich nízke zastúpenie na pozíciách vyššieho a vrcholového manažmentu, čo znamená obmedzené možnosti žien pri spolurozhodovaní. Podľa výsledkov VZPS v roku 2023 z celkového počtu pracujúcich bolo 4,3 % zamestnancov v riadiacej pozícii a v tom podiel žien bol len 1,4 % a podiel mužov 2,9 %.

Vzájomný pomer pohlaví vo vedúcich a riadiacich funkciách je výrazne v prospech mužov, keď na 100 mužov vo vrcholových manažérskych pozíciách pripadla ani nie polovica žien. Z hľadiska dlhodobého vývoja na Slovensku sa rozdiely medzi mužmi a ženami v tomto ukazovateli zlepšujú v prospech žien len nepatrne (tabuľka č. 2).

Tabuľka č. 2: Pracujúci vo vedúcich a riadiacich funkciách

Ukazovateľ		2005	2010	2015	2020	2021	2022	2023
Pracujúci v tis. osobách	spolu	2 357,2	2 357,2	2 433,7	2 365,8	2 317,5	2 315,3	2 329,0
	ženy	1 035,6	1 035,6	1 070,0	1 039,4	1 033,0	1 023,1	1 025,5
	muži	1 321,6	1 321,6	1 363,7	1 326,4	1 284,5	1 292,2	1 303,5
Zákonodarcovia a riadiaci zamestnanci v tis. osobách	spolu	128,4	128,4	134,0	137,1	139,1	120,9	100,8
	ženy	39,7	39,7	39,8	44,2	48,1	37,3	33,4
	muži	88,7	88,7	94,2	92,9	91,0	83,6	67,4
Počet žien na 100 mužov v riadiacich pozíciách		44,8	44,8	42,3	47,6	52,9	44,6	49,6

Zdroj: Štatistický úrad SR, výberové zisťovanie pracovných síl podľa Klasifikácie zamestnaní ISCO-08

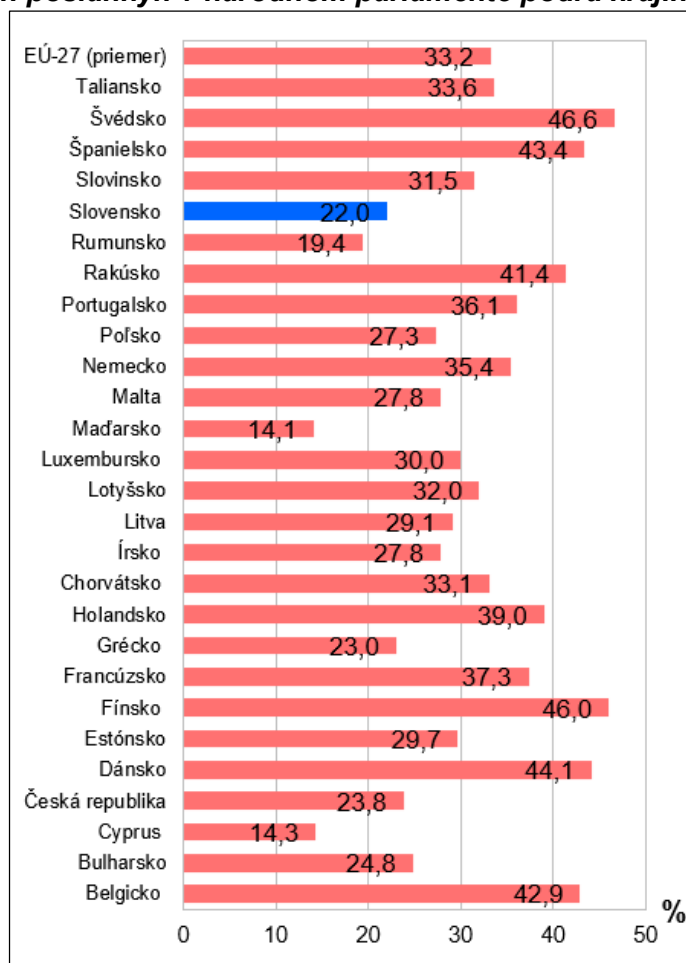
Podobne je to s participáciou žien vo verejnom živote a v politike. Hoci sa demokratické európske krajiny zasadujú za rodovú rovnosť, v politike a verejnom živote na miestnej, vnútroštátnej a európskej úrovni je stále menej žien ako mužov, ako ukazujú najnovšie údaje. Počet poslankyň **Európskeho parlamentu** za posledné desaťročia značne narástol. V období od roku 1952 do prvých priamych eurovolieb v roku 1979 pôsobilo v parlamente len 31 poslankyň. Po prvých priamych voľbách v roku 1979 sa počet žien zvýšil na 15,9 % a narastal po každých nových voľbách, ktoré sa konajú raz za päť rokov. Koncom 2. štvrtroka 2024 bolo v Európskom parlamente 39,3 % poslankyň žien

z celkového počtu zvolených poslancov, za Slovensko 28,6 % z celkového počtu zvolených slovenských poslancov do Európskeho parlamentu.

Nízka účasť žien vo verejnom živote a v politike vedie k tomu, že ženy majú obmedzené možnosti ovplyvňovať strategické rozhodovanie a plánovanie, ktoré má v konečnom dôsledku dosah aj na ich životnú situáciu. Napriek skutočnosti, že záujem o volené funkcie v samosprávnych orgánoch aj v národnom parlamente sa zvyšuje u oboch pohlaví, čo sa prejavilo rastúcim počtom kandidátov vo všetkých typoch volieb, úspešnosť kandidujúcich žien bola oproti mužom nižšia.

Ženy v SR boli najúspešnejšie vo voľbách do Európskeho parlamentu. Počet poslankýň za Slovenskú republiku po posledných voľbách do Európskeho parlamentu (8. júna 2024) vzrástol na 7 (z 15 zvolených poslancov), resp. na 46,7 %. Najmenej úspešné boli ženy vo voľbách na funkciu **predsedu samosprávneho kraja** v SR, kde bola zvolená jediná žena (12,5 %). Do funkcie **poslanca samosprávneho kraja** bolo zvolených 22,8 % žien, čo je o 3,2 p. b. viac ako v predchádzajúcom funkčnom období. Funkciu **starostu obce alebo primátora mesta** vykonáva 26,5 % žien, čím sa ich počet od predchádzajúcich volieb do orgánov samosprávy obcí zvýšil o 30 žien, resp. jeden percentuálny bod.

Graf č. 4: Podiel žien poslankýň v národnom parlamente podľa krajín EÚ (rok 2023)



Zdroj: EIGE – Európsky inštitút pre rodovú rovnosť

Vo voľbách do **Národnej rady SR** (NR SR) v septembri 2023 z celkového počtu kandidátov ženy tvorili 25 %, čo svedčí o ich vyššom záujme oproti predchádzajúcim voľbám z roku 2020 a náraste o 1,9 p. b. Do NR SR bolo zvolených spolu 33 žien (22 %), len o jednu viac než v predchádzajúcich voľbách.

Úspešnosť žien, vyjadrená pomerom počtu zvolených žien k počtu kandidujúcich žien dosiahla 5 % oproti 5,9 % u mužov. Z európskeho pohľadu možno slovenské ženy zaradiť medzi menej úspešné, keďže menšie než štvrtinové zastúpenie v národnom parlamente má iba 7 krajín EÚ27 (graf č. 4).

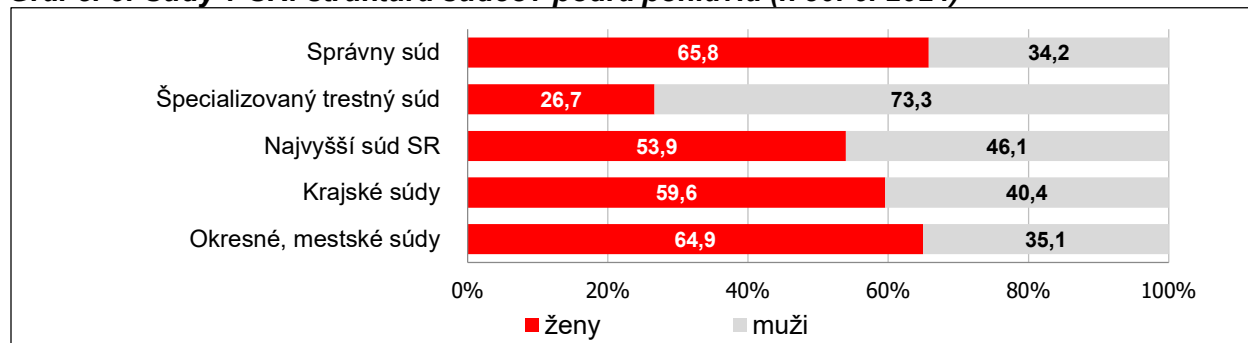
Tabuľka č. 3: Účasť vo verejnom živote a v rozhodovaní

Indikátor	Ženy		Muži		Rozdiel	
	SR	EÚ27	SR	EÚ27	SR	EÚ27
Politická participácia na európskej úrovni (v %, Európsky parlament, 2023, EP/EIGE)	28,6	40,0	71,4	60,0	42,8	20,0
Politická participácia na národnej úrovni (v %, národný parlament, 2023, NR SR/EIGE)	22,0	33,0	78,0	67,0	56,0	34,0
Politická participácia v národných vládach (v %, 2023, vláda SR/EIGE)	14,6	35,2	85,4	64,8	70,8	29,6
Politická participácia na regionálnej úrovni – vo vedení regionálnych zastupiteľstiev (v %, 2023)	12,5	–	87,5	–	75,0	–
Politická participácia na regionálnej úrovni – zastúpenie v krajských/regionálnych zastupiteľstvách (v %, 2023, EIGE)	13,6	35,1	86,4	64,9	72,8	29,8
Politická participácia na úrovni miest a obcí – primátori/primátorky a starostovia/starostky (v %, 2023, EIGE)	27,1	34,5	72,9	65,5	45,8	31,0
Politická participácia – vedenie politických strán s najmenej 5 % miest v národných parlamentoch (v %, 2023, EIGE)	0,0	26,2	100,0	73,8	100,0	47,6
Miera participácie na riadení v najväčších firmách kótovaných na burze (vedúci pracovníci/pracovníčky v %, 2023, EIGE)	2,1	33,0	75,0	67,0	50,0	34,0
Skladba Najvyššieho súdu, členenie podľa pohlavia (v %, 2023, EIGE)	53,3	56,5	46,7	43,5	-6,6	-13,0

Poznámka: rozdiel medzi ženami a mužmi = hodnoty mužov mínus hodnoty žien v percentuálnych bodoch, ak nie je uvedené inak.

Zdroj: Štatistický úrad SR, Eurostat, EIGE

Graf č. 5: Súdny v SR: štruktúra sudcov podľa pohlavia (k 30. 6. 2024)



Zdroj: Štatistický úrad SR, Ministerstvo spravodlivosti SR

Menšinové zastúpenie žien je výrazné aj vo väčšine vrcholových a ústredných orgánov štátnej správy. V súdnictve je výrazná feminizácia (graf č. 5), funkciu sudcu na Slovensku (k 30. 6. 2024) vykonávalo 62,5 % žien a 37,5 % mužov. S výnimkou Špecializovaného trestného súdu SR vo všetkých typoch súdov prevažovalo zastúpenie žien. Čím nižšia je rozhodovacia úroveň súdov, tým je vyšší podiel žien medzi sudcami. Na Najvyššom súde SR ženy predstavovali 53,9 % sudcov, na krajských súdoch 59,6 %, na okresných a mestských súdoch 64,9 %.

5. ZÁVER

Problematika rovnosti medzi ženami a mužmi v rôznych oblastiach života je v centre záujmu najvyšších riadiacich orgánov a inštitúcií. Získané štatistické údaje triedené podľa mužov a žien sú dôležitým východiskom na prijímanie rozhodnutí a ďalších opatrení na zlepšenie života a znižovanie sociálnych a ekonomických rozdielov medzi mužmi a ženami v súčasnej spoločnosti.

Na Slovensku je dlhodobo zaznamenaná výrazná platová nerovnosť medzi mužmi a ženami spojená s množstvom spoločenských, kultúrnych, právnych a ekonomických faktorov, ktoré idú ďaleko nad rámec samotného problému rovnakej odmeny za rovnakú prácu. Pozitívnu správou je skutočnosť, že tieto rozdiely v odmeňovaní medzi mužmi a ženami na Slovensku sa mierne znižujú. Napriek zaznamenanému pokroku v tejto oblasti Slovensko stále patrí ku krajinám Európskej únie s jeho najvyššou hodnotou.

Záujem žien aj mužov o účasť vo voľbách do národného parlamentu, Európskeho parlamentu, orgánov samosprávy obcí a miest, ako aj vo voľbách do orgánov samosprávnych krajov sa s každým volebným obdobím zvyšuje, čo sa prejavuje vyšším počtom kandidátov na volenú funkciu u oboch pohlaví. Miera úspešnosti slovenských žien vo voľbách, vyjadrená pomerom počtu kandidátov k počtu zvolených kandidátov, je vo všeobecnosti nižšia, než u mužov. Napriek tomu, ženy sa neprestávajú snažiť o získanie rovnakej možnosti formovať spoločnosť na rôznych stupňoch riadenia a o rovnomerné rozdelenie moci a vplyvu pri rozhodovaní. Nakoľko menšie než štvrtinové zastúpenie v národnom parlamente má iba 7 krajín Európskej únie, možno slovenské ženy v tomto ich snažení považovať za menej úspešné.

Ing. Emília ČIČVÁKOVÁ

Autorka pracuje na Pracovisku Štatistického úradu SR v Košiciach.

Informácia/Information

DRUHÝ ROČNÍK EURÓPSKEJ SÚŤAŽE V ŠTATISTIKE: SLOVENSKÍ STREDOŠKOLÁCI OPĀŤ EXCELUJÚ

SECOND EDITION OF THE EUROPEAN STATISTICS COMPETITION: SLOVAK HIGH SCHOOL STUDENTS EXCEL ONCE AGAIN

Štatistický úrad Slovenskej republiky úspešne ukončil druhý ročník Európskej súťaže v štatistike 2023/2024, zameranej na stredoškólačkov. Cieľom tejto súťaže je naučiť mladú generáciu rozpoznávať fakty od dezinformácií a využívať oficiálne štatistické údaje. Súťaž sa koná v spolupráci s Eurostatom a pod záštitou Ministerstva školstva, vedy, výskumu a športu SR.

Ciele a význam súťaže

Európska súťaž v štatistike je iniciatívou Eurostatu a národných štatistických úradov EÚ. Zameriava sa na zlepšenie schopností mladých ľudí čítať a interpretovať štatistické informácie a zároveň sa v nich orientovať. Študenti majú možnosť prepojiť interpretáciu dát s kreatívnou prezentáciou alebo videom, čo umožňuje pedagógom využívať alternatívne formy výučby vrátane reálnych údajov z oficiálnych národných a európskych štatistík.

Rekordná účasť a výsledky

Súťaž v školskom roku 2023/2024 zaznamenala výrazné zlepšenie účasti. V porovnaní s rokom 2022/2023 sa do súťaže zapojilo o 50 % viac škôl (129 škôl), o 43 % viac tímov (478 tímov) a o 39 % viac študentov (1 249 študentov). Tento nárast potvrdzuje rastúcu popularitu súťaže. Najväčší nárast počtu prihlásených škôl a tímov zaznamenali Bratislavský, Prešovský a Žilinský kraj. Zároveň sa zvyšuje aj typová rozmanitosť škôl – okrem gymnázií sa zúčastňujú i stredné odborné školy, obchodné akadémie a priemyselné školy.



Zdroj: Eurostat, Štatistický úrad SR

Generálna riaditeľka sekcie diseminácie a komunikácie Štatistického úradu SR, Ing. Nikoleta Šlachtová, zdôrazňuje, že cieľom súťaže je vybaviť mladú generáciu dátovou gramotnosťou, ktorá mladým pomôže identifikovať kvalitné zdroje informácií, rozpoznať dezinformácie a manipulácie, ako aj rozhodovať sa na základe faktov.

Priebeh a kategórie súťaže

Súťaž prebieha online a pozostáva z dvoch kôl – národného a európskeho. Je rozdelená do dvoch kategórií – kategória A (vek 16 až 19 rokov) a kategória B (vek 14 až 16 rokov). Tri najúspešnejšie tímy v každej kategórii národného kola získavajú peňažné poukážky, pričom dva najúspešnejšie tímy z každej kategórie postupujú do európskeho finále.

V národnom kole súťažiaci riešia odborné testy zamerané na štatistické pojmy a produkty Štatistického úradu SR a Eurostatu. Postupujúce tímy potom pripravujú kreatívnu interpretáciu zadaných dát v podobe prezentácie. Testy aj téma prezentácie sa každý rok menia. Priemerné bodové skóre tímov bolo v druhom ročníku o 20 % vyššie v porovnaní s prvým ročníkom.

Vítazi a medzinárodné úspechy

Slovenské tímy dosahujú mimoriadny úspech aj v európskom kole. V druhom ročníku ESC 2023/2024 tím PERO4F z Gymnázia Jozefa Gregora Tajovského v Banskej Bystrici získal 2. miesto v kategórii A a tím SECONDTRY z Gymnázia Ivana Kraska v Rimavskej Sobote získal 2. miesto v kategórii B. Tím TURBO z Gymnázia Považská Bystrica sa umiestnil na 4. mieste v kategórii A.



Vítazi národného kola Európskej súťaže v štatistike s členmi slovenskej odbornej poroty



Tím PERO4F (kategória A) v Madride

Zdroj: Štatistický úrad SR

Inšpiratívny príbeh

Druhý ročník súťaže priniesol aj inšpiratívny príbeh. Gymnazista Pavel z Rimavskej Soboty, ktorý trpí autizmom, získal druhé miesto v európskom kole. Jeho príbeh zaujal i slovenské médiá, čo robí súťaž ešte významnejšou.

Slávnostné odovzdávanie cien

Slávnostné odovzdávanie cien národného kola sa uskutočnilo 19. apríla 2024 v Bratislave a oceňovanie víťazov európskeho kola sa konalo 26. júna 2024 v Madride. Úspešní slovenskí študenti sa zúčastnili na tejto významnej udalosti a prevzali si svoje ocenenia.

Druhý ročník Európskej súťaže v štatistike 2023/2024 potvrdzuje záujem o zvyšovanie dátovej gramotnosti medzi mladými ľuďmi na Slovensku. Tento úspech je povzbudením pre ďalšie ročníky a motiváciou pre študentov, aby sa aktívne zapájali do podobných vzdelávacích aktivít.

Jasmína STAUDER

Autorka je koordinátorka Európskej súťaže v štatistike a hovorkyňa Sčítania obyvateľ'ov, domov a bytov 2021.

Informácia/Information

22. SLOVENSKÁ ŠTATISTICKÁ A DEMOGRAFICKÁ KONFERENCIA

22nd SLOVAK STATISTICS AND DEMOGRAPHIC CONFERENCE

Slovenská štatistická a demografická spoločnosť (SŠDS) v spolupráci so Štatistickým úradom SR (ŠÚ SR) zorganizovala v dňoch 19. – 20. septembra 2024 **22. slovenskú štatistickú a demografickú konferenciu**. V jej tohtoročnom dejisku – v meste Nitra – sa stretlo 76 prihlásených účastníkov. Organizačný výbor pracoval pod vedením riaditeľky pracoviska Štatistického úradu SR v Nitre Ing. Renáty Dušovej a odviezol skvelú prácu pri zabezpečení celej akcie v náročných podmienkach cenovej inflácie.

Program konferencie bol zložený z dvoch tematických častí – štatistickej a demografickej a v štyroch blokoch odznelo 27 príspevkov.

Aktuálny 22. ročník Štatistickej a demografickej konferencie oficiálne otvorila predsedníčka Slovenskej štatistickej a demografickej spoločnosti Iveta Stankovičová a podpredsedníčka Štatistického úradu SR Ľudmila Ivančíková. Vo svojej úvodnej reči

Ľ. Ivančíková upriamila pozornosť na nevyhnutnosť implementácie nových metód, zdrojov a technológií pri rozvoji štatistiky. Súčasne tiež podčiarkla potrebu silnej a trvalej interdisciplinárnej spolupráce s akademickými a výskumnými inštitúciami pri rozvoji novej štatistiky, testovaní nových metód a technológií a podpore inovácií a experimentovania. Prítomní zástupcovia partnerských univerzít v Nitre – František Petrovič, dekan Fakulty prírodných vied a informatiky Univerzity Konštantína Filozofa a Radovan Savov, prodekan Slovenskej poľnohospodárskej univerzity, v tejto súvislosti vyzdvihli dlhoročnú spoluprácu so Štatistickým úradom SR. Zároveň zdôraznili, že absolventi oboch univerzít nachádzajú uplatnenie v praxi, a to aj na pozíciách štatistického úradu.

Po otvorení, úvodných slovách a pozdravných prejavoch vystúpila a Nitriansky kraj v číslach predstavila riaditeľka Pracoviska Štatistického úradu SR v Nitre, Ing. Renáta Dušová.

Prezentácie prvého dňa zahŕňali rôzne oblasti **štatistiky**. Konferencia začala prednáškou pozvaného účastníka Petra Knížata (ŠÚ SR) na tému *Nové dátové zdroje a inovácie v oficiálnych štatistikách*, počas ktorej boli predstavené nové a inovatívne zdroje údajov pre štatistické zisťovania a výsledky inovačných projektov na ŠÚ SR. V príbuznej tematike pokračovali aj ďalšie príspevky od pracovníkov Štatistického úradu SR. Albert Ivančík sa venoval Národnému štatistickému systému (NŠS) a vyhodnoteniu jeho vybraných koordinačných nástrojov. Lucia Vanišová a Filip Lipták prezentovali nový zdroj údajov pre štatistické zisťovania – Štatistický register budov, domov a bytov. Pracovníci oddelenia metodológie ŠÚ SR sa zaoberali využitím transakčných údajov od obchodných reťazcov pre cenové štatistiky (Petra Mazureková, Helena Glaser-Opitzová) a tematike ochrany



Z rokovacej sály; Zdroj: fotografia autorky

dôverných štatistických údajov v populačných gridoch (Boris Frankovič, Arnold Simon). Na záver prvého bloku príspevkov Iveta Fričová (ŠÚ SR) prezentovala ciele a úlohy Agendy 2030 pre Slovensko.

V druhom bloku predniesli príspevky z oblasti štatistiky účastníci konferencie zo slovenských aj českých univerzít. Tomáš Želinský (Ekonomická fakulta Technickej univerzity v Košiciach) prezentoval porovnanie alternatívnych prístupov k odhadu chudoby. Jeho kolegyňa Alena Mojsejová spolu s Miroslavou Pastulovou sa mali venovať príjmovej nerovnosti obyvateľstva a jeho dôvere vo vládu SR a inštitúcie, ale pre chorobu na konferencii nevystúpili. Gábor Szűcs (Fakulta matematiky, fyziky a informatiky Univerzity Komenského v Bratislave) vysvetlil matematické prístupy k modelovaniu procesu volatility vo výnosoch vybraných investičných nástrojov, ktoré skúmali spolu s Adriánou Leginusovou. Tomáš Hlavsa (Provozně ekonomická fakulta, Česká zemědělská univerzita v Prahe) predniesol príspevok o prediktívnom modelovaní životaschopnosti fariem v prírodne znevýhodnených oblastiach ČR. Zuzana Poláková (Fakulta ekonomiky a manažmentu, Slovenská poľnohospodárska univerzita v Nitre) s kolegyňou Miriam Pietrikovou sa vo svojich vystúpeniach venovali vzťahom medzi ekonomickými a demografickými ukazovateľmi a vývoju inflácie a jej zložiek v krajinách Európy. Alžbeta Garajová (Centrum informačných technológií, Univerzita Komenského v Bratislave) predstavila REDCap a SAS EG ako nástroje na tvorbu a spracovanie štatistík v akademickom prostredí. Na záver Emília Čičvákova z Pracoviska ŠÚ SR v Košiciach predniesla zaujímavé číselné fakty o ženách na trhu práce a vo verejnom živote na Slovensku.

Večer počas prvého dňa konferencie sa konal slávnostný raut s panelovou diskusiou a spoločenským programom, na ktorom predsedníčka SŠDS Iveta Stankovičová poďakovala za dlhoročnú prácu v oblasti štatistiky a pomoc pri organizovaní konferencií SŠDS riaditeľom krajských pracovísk ŠÚ SR, pri príležitosti ich odchodu do dôchodku. Odovzdala im ako prejav úcty a vďaky pamätné listy SŠDS. Pamätné listy dostali Ing. Silvia Szabová (riaditeľka pracoviska v Bratislave), Ing. Oľga Chovanová (riaditeľka pracoviska v Žiline), Ing. Ján Cuper (riaditeľ pracoviska v Prešove) a Ing. Anna Janusová (riaditeľka pracoviska v Košiciach) (na priloženej fotografii sprava doľava).



Druhý deň konferencie bol venovaný **demografickej tematike**. S prednáškou *Ako nízko môže klesnúť plodnosť? Zamyslenie nad trendami a scenármi v Európe a vo svete* vystúpil pozvaný hosť Tomáš Sobotka, významný český demograf, ktorý pracuje v Rakúsku vo Viedni ako zástupca riaditeľa Viedenského inštitútu pre demografiu (Rakúska akadémia vied a Wittgensteinovo centrum). Počas prednášky sa účastníci konferencie dozvedeli mnoho zaujímavých faktov o plodnosti žien v krajinách sveta, napr. že najnižšia celková miera plodnosti v roku 2023 bola v Južnej Kórei (len 0,76) a najvyššia v krajinách centrálnej Afriky (Kongo 6,1). V Európe je najnižšia plodnosť v Španielsku (1,1) a Taliansku (1,2) a najvyššia v Bulharsku (1,8). Na Slovensku je to 1,5 a v Česku 1,4.

V treťom bloku príspevkov sa predstavili demografi zo Slovenska. Boris Vaňo (Infostat)

spolu s podpredsedníčkou ŠÚ SR Ľudmilou Ivančíkovou vysvetlili, ako by sa mohla zisťovať populačná klíma na Slovensku. Branislav Bleha (Prírodovedecká fakulta Univerzity Komenského v Bratislave) predniesol demografickú prognózu pre kraje a okresy SR do roku 2050, ktorú vypracoval spolu s kolegami z Infostatu B. Šprochom a B. Vaňom. Branislav Šprocha (Infostat a Slovenská akadémia vied, Bratislava) sa venoval zmenám v sobášnosti a rozvodovosti na Slovensku pred pandémiou ochorenia Covid-19, počas pandémie a po nej. Na záver bloku vystúpil Martin Kočiš (ŠÚ SR, spoluautorka príspevku Veronika Krišková) a prezentoval spôsob ako sa modernizuje demografická štatistika na Slovensku.

V poslednom bloku príspevkov mal vystúpiť známy český demograf Tomáš Kučera (Prírodovedecká fakulta Univerzity Karlovej) a mal hovoriť o predurčení budúceho vývoja obyvateľstva Česka vzhľadom na aktuálnu vekovú a pohlavnú štruktúru. Cieľom príspevku bolo poukázať na nereálnosť niektorých predstáv o ovplyvniteľnosti budúceho populačného vývoja a zároveň prispieť k reálnejším úvahám a zmysluplnejším diskusiám súčasných i budúcich politických reprezentácií. Bohužiaľ pre chorobu sa do Nitry nedostavil. Erika Letrichová (Fakulta matematiky, fyziky a informatiky Univerzity Komenského v Bratislave) sa zaoberala matematickým modelovaním plodnosti v ľudských populáciách a Sofia Karina Trommelová (Prírodovedecká fakulta Univerzity Komenského v Bratislave) hodnotila vplyv tehotenského príspevku na pôrodnosť v SR. Posledný blok uzavreli pracovníci ŠÚ SR. Andrej Chromeček predstavil publikáciu SODB pre najmenších, ktorá patrí medzi veľmi zaujímavé a názorné publikačné výstupy zo sčítania 2021 na Slovensku. Jasmína Stauder spolu s kolegyňou Nikoletou Šlachtovou vypracovali informáciu o druhom ročníku Európskej súťaže v štatistike, v ktorej slovenskí stredoškóľáci excelovali a v závere informovala, že práve rozbiehajú tretí ročník.

Popoludní záverečného dňa konferencie bola zorganizovaná aj dvojhodinová prehliadka historického centra mesta Nitra so sprievodkyňou, na ktorú sa prihlásilo viac ako 25 účastníkov konferencie.

Myslíme si, že 22. slovenská štatistická a demografická konferencia SŠDS v Nitre naplnila ciele a očakávania organizátorov aj účastníkov. Mladí účastníci mali možnosť osobne spoznať odborníkov z radov štatistikov a demografov, ktorých články a publikácie študovali a účastníci skôr narodení mali možnosť sa opäť stretnúť a vymeniť si poznatky a skúsenosti.

Považujeme si za povinnosť poďakovať predsedovi Štatistického úradu SR Ing. Martinovi Nemkymu, MBA a podpredsedníčke PhDr. Ľudmile Ivančíkovej, PhD. za významnú podporu a pomoc pri organizovaní tohto odborného stretnutia. Za skvelú organizáciu celého podujatia ďakujeme organizačnému tímu pod vedením riaditeľky pracoviska ŠÚ SR v Nitre Ing. Renáty Dušovej a jej kolegom a kolegyniam, konkrétne Mgr. Rastislavovi Kvaltýnovi, Mgr. Anne Hricovej, Mgr. Daniele Havlasovej, Mgr. Andrei Horňákovej a Ing. Márii Rybanskej.

Organizačný tím zostavil zborník abstraktov z 22. slovenskej štatistickej a demografickej konferencie SŠDS v Nitre, ktorý je dostupný online na webovej stránke spoločnosti tu:

http://ssds.sk/casopis/konferencie/Zbornik_abstraktov_22_SSK.pdf

**doc. Ing. Iveta STANKOVIČOVÁ, PhD., predsedníčka vedeckého výboru 22SSK
a predsedníčka SŠDS**

Mgr. Silvia HUDECOVÁ, Štatistický úrad SR

Informácia/Information

NOVÝ PREDSEDA ŠTATISTICKÉHO ÚRADU SLOVENSKEJ REPUBLIKY

NEW PRESIDENT OF THE STATISTICAL OFFICE OF THE SLOVAK REPUBLIC

Novým predsedom Štatistického úradu Slovenskej republiky je od februára 2024 **Ing. Martin Nemky, MBA**.



Ing. Martin Nemky, MBA
Zdroj fotografie: ŠÚ SR

Vyštudoval Slovenskú poľnohospodársku univerzitu v Nitre, Fakultu ekonomiky a manažmentu. V roku 2017 získal titul MBA v odbore Agribusiness and Commerce.

Svoju pracovnú kariéru začal ako kontrolór na daňovom úrade v Nitre a na Najvyššom kontrolnom úrade SR. Ako poslanec pôsobil v Nitrianskom samosprávnom kraji a v rokoch 2016 – 2024 v Národnej rade SR. Medzitým zastával funkciu prednostu obvodného úradu v sídle kraja a ako zástupcu primátora mesta Nitra.

Svoje manažérske zručnosti využije v nastavení a implementácii stratégie Štatistického úradu SR v zmysle požiadaviek národných aj európskych používateľov údajov, v nastavení pozície úradu v novom dátovom ekosystéme a v posilnení dôvery úradu.

PRIPRAVUJEME/COMING SOON

Boris FRANKOVIČ

VÝPOČET RÝCHLYCH ODHADOV V ŠTATISTIKE CESTOVNÉHO RUCHU
CALCULATION OF FLASH ESTIMATES IN TOURISM STATISTICS

Martin BOĎA

METODOLOGICKÁ POZNÁMKA K VÝZNAMU A INTERPRETÁCII OKUNOVHO
ZÁKONA
METHODOLOGICAL NOTE ON THE MEANING AND INTERPRETATION OF OKUN'S
LAW

Alena KAŠČÁKOVÁ, Ľudmila IVANČÍKOVÁ, Zuzana RIGOVÁ

AKTÍVNE STARNUTIE V REGIÓNOCH SLOVENSKA
ACTIVE AGEING IN SLOVAK REGIONS

Roman PAVELKA

MOŽNOSTI POUŽITÍ METOD MNOHONÁSOBNÉ IMPUTACE V PROSTŘEDÍ
SYSTÉMU SAS
POSSIBILITIES OF USING MULTIPLE IMPUTATION METHODS IN THE SAS SYSTEM
ENVIRONMENT

* * *

**ONLINE VERZIA ČÍSLA 4/2024 SLOVENSKEJ ŠTATISTIKY A DEMOGRAFIE JE
VEREJNE DOSTUPNÁ na internetovej stránke slovak.statistics.sk a ssad.statistics.sk od
15. OKTÓBRA 2024.**

**THE ONLINE VERSION OF THE JOURNAL SLOVAK STATISTICS AND
DEMOGRAPHY No 4 (2024) IS PUBLICLY BE AVAILABLE at the website
slovak.statistics.sk and ssad.statistics.sk from **OCTOBER 15, 2024.****

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 poskytnite 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 nájdú autori na ssad.statistics.sk.

Rozsah vedeckých článkov je okolo 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.

Scope of a scientific article is about 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.

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 súčasť demografickej štatistiky.

Vydáva:

Štatistický úrad Slovenskej republiky
Slovenská štatistická a demografická spoločnosť

Identifikačné čísla vydavateľov:

IČO 00166197 / 00178764

Vychádza:

Štyrikrát ročne

Dátum vydania:

15. október 2024

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

Evidenčné číslo/Evidence number 272/08

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 Slovak Republic
Slovak Statistical and Demographic Society

Companies registration numbers:

00166197 / 00178764

Published:

Four times a year

Date of issue:

15th October 2024

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