

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

2/2024  
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](http://ssad.statistics.sk) a na [slovak.statistics.sk](http://slovak.statistics.sk). Názory autorov článkov sa nemusia zhodovať s názormi vydavateľa.

### Zahraniční poradcovia/Foreign Consultants

#### **Gabriela Czanner**

University of Liverpool  
Veľká Británia/United Kingdom

#### **Jitka Langhamrová**

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

#### **Estefanía Mourelle Espasandín**

Universidade da Coruña  
Španielsko/Spain

#### **Michaela Potančoková**

Joint Research Centre,  
European Commission  
Taliansko/Italy

#### **Hana Řezanková**

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

#### **Milan Stehlík**

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

### Výkonná redaktorka/Executive Editor

Silvia Hudecová

### Jazykové redaktorky/Language Editors

#### **Slovenský jazyk/Slovak Language**

Silvia Duchková

#### **Anglický jazyk/English Language**

Andrea Okenková

### Adresa redakcie/Address of Editorial Office

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

## 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](http://ssad.statistics.sk) and [slovak.statistics.sk](http://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

### E-mailová adresa/E-mail address

[SSaD@statistics.sk](mailto:SSaD@statistics.sk)

[ssad.statistics.sk](http://ssad.statistics.sk)  
[www.statistics.sk](http://www.statistics.sk)

## OBSAH/CONTENTS

### I. VEDECKÉ ČLÁNKY/SCIENTIFIC ARTICLES

**Roman PAVELKA** 3  
STATISTICKÁ ANALÝZA CHYBĚJÍCÍCH DAT  
STATISTICAL ANALYSIS OF MISSING DATA

**Boris VAŇO** 26  
VSTUP SLOVENSKÝCH ŽIEN DO REPRODUKCIE  
THE SLOVAK WOMEN'S ENTRY INTO REPRODUCTION

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

**Emília MLADÁ** 39  
VYUŽITIE VYMIENANÝCH MIKROÚDAJOV NA SPRESNENIE ŠTRUKTÚRY  
ÚDAJOV ŠTATISTIKY ZAHRANIČNÉHO OBCHODU S TOVAROM  
USE OF EXCHANGED MICRODATA TO REFINE THE DATA STRUCTURE  
OF FOREIGN TRADE IN GOODS STATISTICS  
Informatívny článok/Informative article

**Mikuláš CÁR** 43  
INTERAKTÍVNA PUBLIKÁCIA O BÝVANÍ V EURÓPE  
INTERACTIVE PUBLICATION ABOUT HOUSING IN EUROPE  
Informácia/Information

**Mária VOJTKOVÁ** 49  
ZA MARTINOU LUBYOVOU  
IN MEMORY OF MARTINA LUBYOVÁ  
Nekrológ/Necrology

**III. PRIPRAVUJEME/COMING SOON** 51



**Roman PAVELKA**  
**Štatistický úrad Slovenskej republiky**

## **STATISTICKÁ ANALÝZA CHYBĚJÍCÍCH DAT**

## **STATISTICAL ANALYSIS OF MISSING DATA**

### **ABSTRAKT**

Standardní statistické metody byly vyvinuty pro analýzu souborů dat v maticovém uspořádání. Řádky datové matice tradičně reprezentují jednotky, označované také jako případy, pozorování nebo předměty – v závislosti na kontextu. Měřené, resp. zjišťované proměnné nebo také charakteristiky pro každou sledovanou jednotku představují sloupce matice dat. Údaje v datech matice jsou téměř vždy reálná čísla u spojitých proměnných, jako je věk anebo příjem, tržby, nebo představují kategoriální odpovědi, které mohou být uspořádané (např. velikostní kategorie, úroveň vzdělání) nebo neuspořádané (nominální), jako je například odvětví ekonomické činnosti, pohlaví, rasa apod. V praxi výběrových zjišťování se však často objevují datové matice pozorovaných hodnot, ve kterých hodnoty některých charakteristik nejsou zaznamenány a jsou chybějící. Například chybějící hodnoty tržeb, obratu anebo jiných ekonomických ukazatelů v podnikových zjišťováních nebo odmítnutí poskytnutí hodnoty příjmu u respondentů v šetření domácností. Příspěvek se zabývá statistickou analýzou takových datových matic, ve kterých hodnoty jedné nebo více proměnných nejsou kompletně vyplněny.

### **ABSTRACT**

Standard statistical methods have been developed for the analysis of data sets in a matrix arrangement. The rows of a data matrix traditionally represent units, also referred to as cases, observations, or subjects - depending on the context. The measured or the surveyed variables or characteristics for each monitored unit represent the columns of the data matrix. Data in matrix data are almost always real numbers for continuous variables such as age or income, turnover, or represent categorical responses that can be ordered (e.g. size category, level of education) or unordered (nominal) such as a sector of economic activity, gender, race, etc. In the practice of sample surveys, however, data matrices of the observed values often appear in which the values of some characteristics are not recorded and are missing. For example, the turnover's missing values, turnover and/or other economic indicators in business surveys or refusal to provide income values for respondents in household surveys. The paper deals with the statistical analysis of such data matrices in which the values of one or more variables are not completed in full.

### **KLÍČOVÁ SLOVA**

mechanismus chybění, neúplná data, statistická inference, statistické modelování

### **KEY WORDS**

missing data mechanism, incomplete data, statistical inference, statistical modelling

## **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 a periodicitě 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 nechtějí. 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?

Za chybějící údaje jsou považovány takové údaje, které v datové matici chybí u některých (nikoliv u všech) proměnných a u některých (nikoliv u všech) respondentů (případů, resp. zpravodajských jednotek). Tyto chybějící údaje v datové matici jsou problematické především z toho důvodu, že většina současných analytických programů a nástrojů předpokládá datovou matici úplnou. Pokud tomu tak není, dochází při analýzách nejčastěji k ignorování případů, jimž některé sledované hodnoty chybí. Proto může být tento postup zpracování dat nevhodný a výsledné statistické úsudky mohou být více či méně chybné.

Jedním z problémů neúplných dat je skutečnost, že mnohdy není znám důvod chybějících pozorování. Není jasné, zda mechanismus vzniku chybějících dat<sup>1</sup> je zcela náhodný, nebo zda tento mechanismus závisí na pozorovaných či dokonce na nepozorovaných hodnotách [11]. Cílem tohoto příspěvku je provést statistickou analýzu chybějících dat, a to zejména s důrazem na popis a modelování mechanismu chybějících dat. Po nezbytném teoretickém úvodu bude na několika příkladech umělých dat realizováno modelování tohoto mechanismu a načrtnuty metody správného postupu při statistických analýzách neúplných dat.

## 2. DEFINICE MECHANISMU CHYBĚNÍ DAT A PŘIDRUŽENÝCH VELIČIN

Mechanismus chybějících dat je obvykle definován jako statistický vztah mezi subjekty, proměnnými a pravděpodobnostmi chybějících dat [3]. Tento model umožňuje vypočítat pravděpodobnost chybějícího údaje pro každý zpravodajský subjekt a analyzovanou proměnnou. Příkladem takového modelu pro chybějící data je pravidlo, že „každá zpravodajská jednotka s pravděpodobností 20 % bude mít ve sledované hodnotě proměnné  $Y$  údaj chybějící“. Statistické vyjádření modelu spočívá ve stanovení pravděpodobnosti pro chybějící údaje  $P(M = 1) = 0,2$ , kde  $M$  je náhodná indikátorová proměnná  $M = 1$  indikující chybějící hodnotu v proměnné  $Y$  a  $M = 0$  indikující pozorovanou hodnotu. Statistický model o chybějících datech předpokládá, že pravděpodobnost, že chybí údaj u jednoho subjektu (pozorování), je nezávislá na pravděpodobnosti, že bude chybět údaj v proměnné u subjektu (pozorování) jiného. Předpoklad nezávislosti je společný předpoklad při generování chybějících dat [5]. Zjištěný výběrový soubor může obsahovat i více mechanismů generování chybějících dat, a to jak pro každou proměnnou, tak i pro více proměnných.

Stejně jako jiné druhy statistických modelů má mechanismus chybějících dat jeden nebo více parametrů. Tyto parametry jsou spojeny s rozdělením indikátoru chybějících dat  $M$ . Pro úspěšné modelování chybějících dat je potřebné specifikovat tyto parametry spojené s modelem neúplných dat. Jelikož se tyto parametry modelu neúplných dat týkají populace, parametry je možné pouze odhadovat. Ačkoli se střední

<sup>1</sup> Mechanismus chybějících dat jako proces způsobující chybění v datech byl pro další analýzy dat poprvé formálně definován v práci [11].

hodnota odhadovaných hodnot parametru během opakovaných výběrů z populace rovná skutečné hodnotě parametru, v konkrétních výběrových souborech se odhadovaný parametr obvykle liší od skutečné hodnoty parametru. Znalost pravidla statistického modelu pro chybějící data usnadňuje datovým analytikům zjistit mnoho vlastností vytvářených chybějících dat.

V reálném výzkumu existuje množství důvodů, proč mohou údaje chybět (zabývají se jimi například v práci [7]). Tyto důvody zároveň ovlivňují rozložení chybějících dat v datové matici. Aby bylo možné ze získaných dat vyvodit správné statistické závěry, i za podmínek neodpovědí, je nutné v rámci standardního statistického usuzování zahrnout do statistické inference i vliv zjištěných neodpovědí. Rozdělení chybějících údajů je přitom jedním ze zásadních předpokladů při rozhodování o způsobu práce s těmito daty. Pro zkoumání charakteru rozdělení chybějících dat v datovém souboru proto byly podle [11] zavedeny přidružené indikátorové proměnné, které umožňují přiblížit důvody chybění dat.

### 3. KLASIFIKACE MECHANISMU CHYBĚNÍ DAT

Mechanismus vzniku chybějících údajů, kterým se pro účely analýzy rozumí jejich vztah k hodnotám dalších proměnných v datovém souboru, je pro nalezení adekvátního postupu v procesu statistické inference podstatný. Vysvětlení vzniku chybějících údajů a argumenty pro nakládání s datovými soubory obsahující chybějící údaje je třeba hledat už v procesu sběru dat. Na základě klasifikace uvedené v práci [11] nebo [2] podle mechanismu vzniku mohou pak chybějící údaje být [10]:

- **zcela náhodné** (Missing Completely at Random – MCAR), kdy rozdělení náhodné veličiny  $M$  nezávisí na  $Y$ , neboli pravděpodobnost, že údaj bude u jednotky chybět, nezávisí na hodnotách pozorované či chybějící náhodné veličiny  $Y$  (například chybějící údaje o příjmu nezávisí na tom, zda jde o příjmy malé či velké, příjmy mužů či žen atd.). Jedná se o nejvíce žádoucí případ, který nezpůsobuje zkreslení prováděných odhadů;
- **náhodné** (Missing At Random – MAR), kdy rozdělení náhodné veličiny  $M$  závisí pouze na zjištěných (pozorovaných) hodnotách náhodné veličiny  $Y$  (například chybějí spíše údaje o příjmu mužů, nezávisle na tom, zda jde o příjmy menší či větší). Jedná se o případ chybění, které je také žádoucí. MCAR je zřejmě zvláštním případem MAR;
- **nenáhodné** (Not Missing At Random – NMAR), kdy rozdělení náhodné veličiny  $M$  závisí na chybějících hodnotách sledované  $Y$ . Například pokud chybějí převážně údaje o příjmu mužů, které jsou spíše vyšší. Jedná se nejproblematictější případ.

Formálně se podle [9] výše uvedená klasifikace mechanismu chybění dat dá vyjádřit pomocí podmíněného rozdělení pravděpodobnosti  $P(\mathbf{M}|y)$  indikátoru chybějících dat (náhodné veličiny)  $\mathbf{M}$  za podmínky realizace sledované náhodné veličiny  $Y$ , kde  $Y = (Y_1, \dots, Y_p)^T$  reprezentuje náhodný vektor  $p$  sledovaných proměnných a  $y = (y_1, \dots, y_p)^T$  představuje realizaci  $Y$ . Pro definici jednotlivých typů mechanismů neúplných dat se dále bude předpokládat, že chybějící hodnoty obsahuje pouze náhodná proměnná  $Y_1$ . Náhodný indikátor neúplných dat  $\mathbf{M}$  s hodnotou 1, tj.  $\mathbf{M} = 1$ , bude proto indikovat chybějící hodnoty v proměnné  $Y_1$ .

Zcela náhodný mechanismus chybění dat (anglická zkratka MCAR) vzniká tehdy, kdy rozdělení pravděpodobnosti  $P(\mathbf{M}|y)$  indikátorové veličiny  $\mathbf{M}$  nezávisí na realizovaných hodnotách pozorované či chybějící veličině  $Y$ , tj.

$$P(\mathbf{M} = 1|y) = P(\mathbf{M} = 1) \text{ a } P(\mathbf{M} = 0|y) = P(\mathbf{M} = 0) \quad (1)$$

Předpoklad toho, že data chybějí zcela náhodně (MCAR), patří k těm nejpřísnějším. K definování zbývajících typů mechanismu chybění je nutné rozdělit realizaci sledované náhodné proměnné  $Y_1$  na hodnoty pozorované  $(y_{obs})^T$  a hodnoty nepozorované (chybějící)  $(y_{miss})^T$ , tj.  $y = (y_{obs}, y_{miss})^T$ . Jelikož podle předpokladu chybějící hodnoty obsahuje pouze náhodná proměnná  $Y_1$ , pak platí, že  $y_{miss} = y_1$  a  $y_{obs} = (y_2, \dots, y_p)^T$ .

Náhodné rozložení chybějících dat (MAR) je méně striktním předpokladem. Rozdělení chybějících dat reprezentovaném rozdělením indikátorové veličiny  $\mathbf{M}$  – a i když jsou data rozdělena náhodně – již není nezávislé na sledovaných datech. Pro rozdělení indikátorové veličiny  $\mathbf{M}$  platí, že závisí na pozorovaných hodnotách  $(y_{obs})^T$  náhodné proměnné  $Y$ , ale nezávisí na chybějících hodnotách  $(y_{miss})^T$  této náhodné proměnné tj. pro náhodné rozložení chybějících dat (MAR) platí:

$$P(\mathbf{M} = 1|(y_{obs}, y_{miss})^T) = P(\mathbf{M} = 1|(y_{obs})^T) \text{ a} \quad (2)$$

$$P(\mathbf{M} = 0|(y_{obs}, y_{miss})^T) = P(\mathbf{M} = 0|(y_{obs})^T)$$

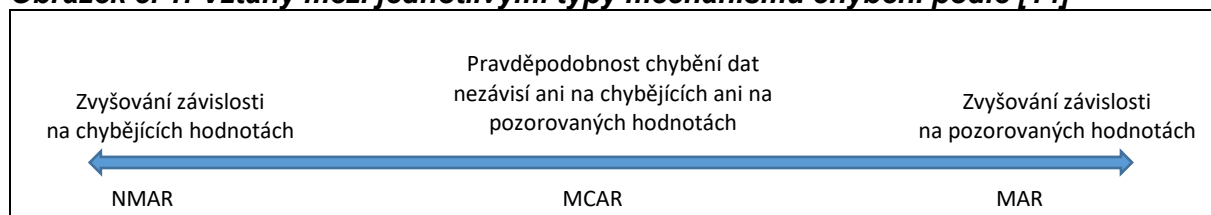
Nenáhodné rozdělení chybějících dat (NMAR) vzniká tehdy, když rozdělení indikátorové proměnné závisí nejen na pozorovaných hodnotách  $Y$ , ale i na hodnotách nepozorovaných (chybějících). Pro náhodné rozdělení indikátoru  $\mathbf{M}$  proto platí:

$$P(\mathbf{M} = 1|(y_{obs}, y_{miss})^T) \text{ a} \quad (3)$$

$$P(\mathbf{M} = 0|(y_{obs}, y_{miss})^T).$$

Z výše uvedených definic jednotlivých typů mechanismu chybění vyplývá, že zcela náhodný mechanismus chybění dat lze být považován za speciální případ mechanismů chybění náhodného a nenáhodného. Jsou-li náhodně neúplná data nezávislá na pozorovaných hodnotách  $(y_{obs})^T$ , tj. závislost mechanismu chybění na pozorovaných hodnotách je nulová, jsou data zcela náhodně chybějící (MCAR). Jsou-li neúplná data závislá na chybějících hodnotách  $(y_{miss})^T$ , tj. závislost mechanismu chybění na nepozorovaných hodnotách je nenulová, jsou data náhodně chybějící (NMAR). Vztahy mezi jednotlivými typy mechanismy chybění pro spojitá rozdělení ilustruje obrázek č. 1.

**Obrázek č. 1: Vztahy mezi jednotlivými typy mechanismu chybění podle [14]**



**Zdroj: vlastní zpracování podle [12]**



#### 4. IGNOROVATELNÝ A NEIGNOROVATELNÝ MECHANISMUS CHYBĚNÍ DAT

Ignorovatelná data jsou typy chybějících dat, které lze efektivně zpracovat moderními technikami pro odhady za podmínek chybějících dat, jako jsou expektační-maximalizační algoritmy, metody maximálně-věrohodné odhadů a metody mnohonásobné imputace. Chybějící data musí splňovat dvě podmínky, aby se stala ignorovatelnými chybějícími daty:

- chybějící údaje jsou buď chybějící podle mechanismu MCAR nebo MAR a
- parametry spojené s konkrétním pravidlem pro chybějící data se liší od parametrů spojených s distribucí proměnných v datovém souboru [11].

Druhá podmínka znamená, že parametry související s rozdělením indikátoru chybění  $M$  se liší od parametrů spojených s pravděpodobnostním rozdělením  $Y$ . Pro pochopení důvodu, proč jsou tyto podmínky potřebné, necht'  $\theta$  resp.  $\varphi$  označují parametry spojené s rozdělením  $Y$ , resp.  $M$ , a necht'  $f(y, m; \theta, \varphi)$  označuje sdruženou hustotu pravděpodobnosti rozdělení  $M$  a  $Y$ . Vzhledem k tomu, že  $\theta$  resp.  $\varphi$  jsou parametry vzájemně odlišné, za podmínek neúplných dat, pravděpodobnost pozorovaných dat lze získat prostřednictvím okrajové (marginální) hustoty  $(y_{obs})^T$  takto:

$$f(y_{obs}, m, \theta, \varphi) = \int f(y_{obs}, y_{miss}; \theta) f(m|y_{obs}, y_{miss}; \varphi) dy_{miss}. \quad (4)$$

Jsou-li neúplná data zcela náhodně chybějící (MCAR), potom:

$$f(m|y_{obs}, y_{miss}; \varphi) = f(m; \varphi). \quad (5)$$

Jsou-li neúplná data náhodně chybějící (MAR), potom:

$$f(m|y_{obs}, y_{miss}; \varphi) = f(m|y_{obs}; \varphi). \quad (6)$$

Jelikož ani  $f(m; \varphi)$ , ani  $f(m|y_{obs}; \varphi)$  nezahrnují  $y_{miss}$ , lze hustoty  $f(m; \varphi)$  i  $f(m|y_{obs}; \varphi)$  vyjmout před integrál (vzorec 4) a pro maximálně-věrohodné odhady je postačující integrál  $\int f(y_{obs}, y_{miss}; \theta) dy_{miss}$  k odhadu parametru  $\theta$  maximalizovat jen a pouze vzhledem k parametru  $\theta$ . I v případě, že neúplná data typu MAR nesplňují druhou podmínku o různých parametrech  $\theta$  a  $\varphi$ , statistické metody předpokládají ignorovatelnost mechanismu chybění (i když ne optimální, ale stále vyhovující pro nevychýlené odhady). V praxi jsou proto mechanismy chybění zcela náhodné (MCAR) a náhodné (MAR) považovány za ignorovatelné a mechanismus nenáhodných chybění (NMAR) implikuje mechanismus chybění neignorovatelný.

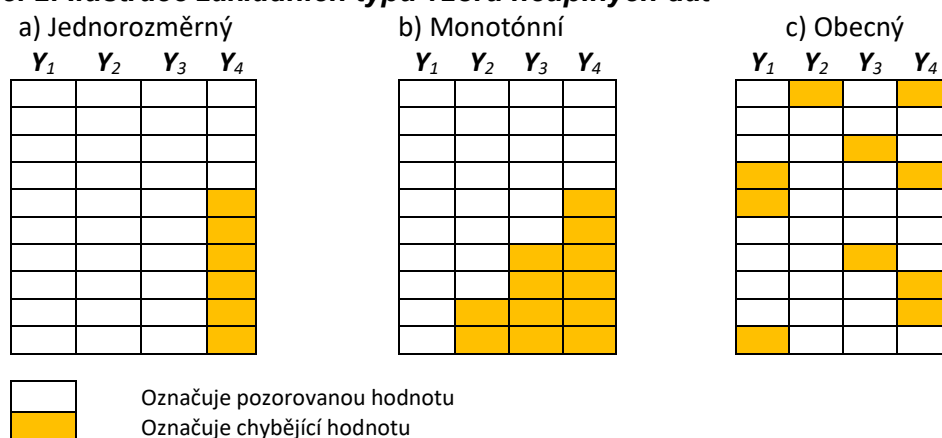
#### 5. VZORY CHYBĚJÍCÍCH DAT

Vzor chybějících dat<sup>2</sup> se týká uspořádání pozorovaných a chybějících hodnot v souboru dat [4]. Často se zaměřuje s mechanismem neúplných dat (např. [8]). Rozdíl je v tom, že specifický mechanismus chybění dat je datové pravidlo v neúplných datech, které popisuje vztah mezi subjekty (zpravodajskými jednotkami) a pravděpodobnostmi chybějících dat. Na druhou stranu specifický chybějící datový vzor je datová konfigurace, která popisuje umístění chybějících hodnot v datech.

<sup>2</sup> Vzor chybějících dat je překlad anglického originálu *Missing Data Pattern*.

Obecně existují tři druhy chybějících datových vzorů. Jednorozměrný vzor vzniká, když chybí hodnoty u jedné proměnné nebo skupiny proměnných, která je buď zcela pozorována, nebo zcela chybí pro každý případ (zpravodajskou jednotku), ale všechny ostatní proměnné jsou zcela pozorovány [12] (viz obrázek č. 2a). Jednorozměrný vzor má nejnižší počet chybějících datových vzorů; jinými slovy, má dva vzory chybějících dat; jeden vzor, kde subjekty mají úplná data, a druhý vzor, kde subjektům více či méně chybí data. Dalším typem chybějícího vzoru je monotónní vzor neúplných dat (např. [12]). V monotónním chybějícím vzoru lze skupinu proměnných  $(Y_1, \dots, Y_p)^T$  seřadit tak, že pokud chybí hodnota  $Y_j$  u  $j$ -té zpravodajské jednotky, pak hodnoty  $(Y_{j+1}, \dots, Y_p)^T$  také chybí (viz obrázek č. 2b). Na jednorozměrný vzor lze pohlížet jako na zvláštní případ monotónního vzoru. A konečně, obecný vzor chybějících dat se vytvoří, když může chybět skupina proměnných pro jakýkoli zpravodajský subjekt. Toto vytváří datový soubor s chybějícími hodnotami rozptýlenými v datové matici náhodným způsobem [4] (viz obrázek č. 2c). Uvedené vzory chybění dat jsou ilustrovány na obrázku č. 2.

**Obrázek č. 2: Ilustrace základních typů vzorů neúplných dat**



**Zdroj: vlastní zpracování podle [12]**

Přestože datový vzor neúplných dat a mechanismus neúplných dat mají odlišný význam, vzájemně se ovlivňují. Vzhledem ke konkrétnímu pravidlu pro chybějící data s určitým typem mechanismu chybějících dat bude určen počet a typ vzoru neúplných dat. Má-li například datová množina proměnné  $(Y_1, \dots, Y_p)^T$  a pokud pravidlem chybějících dat je, že každý subjekt má pravděpodobnost 20 % chybějících hodnot v proměnné  $Y_1$ , pak je vzor neúplných dat jednorozměrný, což se rovná dvěma vzorům chybění.

I když vzory neúplných dat nemají přímý vliv na mechanismus neúplných dat, je znalost vzoru neúplných dat pro volbu efektivní metody odhadů parametrů velmi důležitá [11].

## 6. STATISTICKÉ MODELY CHYBĚJÍCÍCH DAT

### Statistické modely zcela náhodného mechanismu chybění (MCAR)

Zcela náhodný mechanismus chybění je charakterizován pravděpodobnostmi, že u každé zpravodajské jednotky budou data neúplná v jedné anebo více proměnných. Pravděpodobnost chybějícího údaje je hodnota parametru označovaná jako  $\pi$ .

Hodnota parametru má vliv na očekávané procento chybných údajů a očekávaný počet vzorů chybní výběrového souboru.

U dat s mechanismem chybní typu MCAR s jednorozměrným vzorem neúplných dat platí pro chybní údaje pravidlo, že každému zpravodajskému subjektu je přiřazena pravděpodobnost neúplných údajů v proměnné (proměnných). Pro neúplná data tedy platí, že pravděpodobnost chybní je  $P(M = 1) = \pi$ , kde  $M$  je indikátor chybných dat a  $\pi$  je parametr vyjadřující hodnotu pravděpodobnosti chybných údajů. Na základě znalosti tohoto pravděpodobnostního pravidla mohou výzkumníci určit různé vlastnosti spojené s údaji MCAR, včetně očekávaného procenta chybných hodnot a očekávaného počtu chybných vzorů neúplných dat.

K pochopení, jak zcela náhodný mechanismus neúplných dat ovlivňuje vlastnosti dat ve výběrovém souboru, je zavedena náhodná proměnná  $K$  udávající počet zpravodajských jednotek (případů) s chybnými údaji. Za předpokladu nezávislosti šance chybní údajů v proměnné jedné zpravodajské jednotky na šanci chybní u ostatních jednotek, náhodná proměnná  $K$  podléhá binomickému rozdělení  $K \sim Bin(n, \pi)$ , kde  $n$  je celkový počet zpravodajských jednotek ve výběrovém souboru a  $0 \leq \pi \leq 1$ . Jelikož  $E(K) = n\pi$  a  $Var(K) = n\pi(1 - \pi)$ , potom očekávané procento chybných hodnot je:

$$E(\Pi) = E\left(\frac{K}{n}\right) = \frac{1}{n}E(K) = \pi, \quad (7)$$

kde  $\Pi = \frac{K}{n}$  je náhodná proměnná popisující odhadované procento chybných hodnot ve výběrovém souboru. Rozptyl odhadu procenta chybných údajů je:

$$Var(\Pi) = Var\left(\frac{K}{n}\right) = \frac{1}{n^2}Var(K) = \frac{\pi(1-\pi)}{n}. \quad (8)$$

Rozptyl odhadu procenta podle vzorce (8) vyjadřuje skutečnost, že se ve výběrovém souboru nemusí objevit přesná hodnota očekávaného procenta chybných údajů.

Na základě pravděpodobnosti neúplných dat  $\pi$  je možné ve výběrovém souboru také odvodit očekávaný počet vzorů chybní. U dat s mechanismem chybní typu MCAR s jednorozměrným vzorem neúplných dat platí, že jeden vzor zahrnuje zpravodajské jednotky s úplnými daty; vzor druhý zahrnuje zpravodajské jednotky s chybnými hodnotami. Nechť  $I_j$ , pro  $j \in \{1, 2\}$ , je indikační proměnná události, že  $j$ -tý vzor neúplných dat platí alespoň u jedné zpravodajské jednotky z výběrového souboru. Pravděpodobnost, že vzor neúplných dat pro  $j = 1$  nastal alespoň u jedné zpravodajské jednotky, je  $P(I_1 = 1) = E(I_1) = 1 - \pi^n$ . Pravděpodobnost, že vzor neúplných dat pro  $j = 2$  nastal alespoň u jedné zpravodajské jednotky, je  $P(I_2 = 1) = E(I_2) = 1 - (1 - \pi)^n$ . Nechť  $D$  označuje počet odlišných vzorů neúplných dat, tj.  $D = \sum_{j=1}^2 I_j$ . Očekávaný počet odlišných chybných datových vzorů potom je:

$$E(D) = E\left(\sum_{j=1}^2 I_j\right) = \sum_{j=1}^2 E(I_j) = 1 - \pi^n + 1 - (1 - \pi)^n = 2 - \pi^n - (1 - \pi)^n, \quad (9)$$

Se zvyšováním velikosti výběrového souboru  $n$  očekávaný počet vzorů neúplných dat konverguje hodnotě 2 (a to je maximální počet vzorů neúplných dat pro daný mechanismus chybní).

Objevují-li se chybějící hodnoty ve více než 1 proměnné  $Y_1, \dots, Y_l$ , pro  $i \in \{1, \dots, l\}$ , potom lze očekávat při stávajícím mechanismu chybění celkem  $l$  parametrů,  $\pi_1, \dots, \pi_l$ , popisující mechanismů neúplných dat pro  $l$  proměnných. Pro každou proměnnou s neúplnými daty  $Y_i$ , je očekávané procento chybějících hodnot ve výběrovém souboru a jeho rozptyl dané rovnicemi (7) a (8). Daný mechanismus chybění, kde vzor neúplných dat obsahuje až  $l$  proměnných s chybnými údaji, může generovat až  $m = 2^l$  vzorů neúplných dat. Očekávaný počet různých vzorů neúplných dat ve výběrovém souboru s  $l$  proměnnými s chybějícími údaji je definován rovnicí:

$$E(D) = m - \sum_{j=1}^m (1 - \eta_j), \quad (10)$$

kde  $\eta_1, \dots, \eta_m$  jsou odpovídající pravděpodobnosti pro  $m$  vzorů neúplných dat. U zcela náhodného mechanismu chybění (MCAR) pravděpodobnosti pro vytvoření vzorů neúplných dat, tj.  $\eta_1, \dots, \eta_m$ , závisí jen a pouze na pravděpodobnostech neúplných údajů, které se objeví v proměnných  $Y_1, \dots, Y_l$ , pravděpodobnosti  $\pi_1, \dots, \pi_l$ .

### Statistické modely náhodného mechanismu chybění (MAR)

U výběrových souborů s daty generovanými náhodným mechanismem chybění (MAR) závisí pravděpodobnost, že daná jednotka bude mít chybějící hodnotu, na pozorovaných hodnotách jiných proměnných. To znamená, že pravděpodobnost chybějících hodnot lze předpovídat z pozorovaných hodnot jiných proměnných. Proměnnou, která může předpovídat pravděpodobnost chybějících hodnot, je proto možné nazývat jako prediktor chybějících dat. Prediktorem chybějících dat může být jedna proměnná datový soubor nebo to může být nová proměnná, která je lineární kombinací několika proměnných v datovém souboru.

Modelování neúplných dat za podmínek náhodného mechanismu chybění MAR je složitější než modelování dat MCAR ze dvou důvodů:

- Pravděpodobnostní pravidla pro chybějící data jsou složitější než pravidla pro data zcela náhodná a dají se rozdělit do několika kategorií:
  - s jednoduchou rozhodovací (diskriminační) hodnotou,
  - s vícenásobnými rozhodovacími (diskriminačními) hodnotami,
  - percentilové rozhodovací hodnoty,
  - založené na logistické regresi.
- Mezi indikátorem chybějících dat a prediktorem chybějících dat se může měnit síla a tvar závislosti. Síla závislosti může být slabá nebo silná; tvar závislosti může být lineární nebo nelineární.

### Náhodný mechanismus chybění MAR s jednoduchou rozhodovací hodnotou

Pravděpodobnostní pravidla pro chybějící data spojená s jedinou mezní hodnotou zahrnují určení této mezní hodnoty pro každý prediktor chybějících dat. Pro neúplná data s náhodným mechanismem chybění MAR je  $Y_1$  proměnná s chybějícími daty a  $Y_2$  je prediktor chybějících dat s hraničním bodem  $a$ . V tomto případě mají data jednorozměrný vzor chybění. Pravděpodobnostní pravidlo pro neúplná data je: pokud má jednotka hodnotu proměnné  $Y_2 \geq a$ , pak pravděpodobnost, že chybí hodnota  $Y_1$ , je  $\pi_1$ ; pokud má jednotka hodnotu proměnné  $Y_2 < a$ , pak jeho pravděpodobnost, že chybí hodnota  $Y_1$ , je  $\pi_2$ . Formálně lze toto pravděpodobnostní pravidlo definovat takto: nechť  $M$  je indikátor chybějících dat pro chybějící data v proměnné  $Y_1$  a  $U$  je indikátor označující, zda proměnná  $Y_2$  nabývá hodnoty  $a$  a vyšší (tj.  $U = 1$ , když  $Y_2 \geq a$  a  $U = 0$ ,

když  $Y_2 < a$ ). Indikátor  $U$  je náhodná proměnná, také prediktor chybějících údajů, a je lineární funkcí náhodné proměnné  $Y_2$ . Rovnice pravděpodobnostního pravidla pro mechanismus chybění MAR lze zapsat ve tvaru  $P(M = 1|U = 1) = \pi_1$  a  $P(M = 1|U = 0) = \pi_2$ . Pravděpodobnostní pravidlo chybějících údajů zahrnuje pouze dva indikátory  $M$  a  $U$ , a proto jej lze nejlépe ilustrovat pomocí kontingenční tabulky pro tyto dva indikátory. Tato kontingenční tabulka je uvedena na obrázku č. 3.

**Obrázek č. 3: Kontingenční tabulka pro náhodný mechanismus neúplných dat s jednoduchou diskriminační hodnotou**

U	M	
	1	0
1	$P(M = 1 U = 1)P(U = 1) = \pi_1\pi_0$	$P(M = 0 U = 1)P(U = 1) = (1 - \pi_1)\pi_0$
0	$P(M = 1 U = 0)P(U = 0) = \pi_2(1 - \pi_0)$	$P(M = 0 U = 0)P(U = 0) = (1 - \pi_2)(1 - \pi_0)$

$M$  je indikátorová proměnná označující, kdy je hodnota proměnné  $Y_1$  chybějící:  $M = 1$ , když je hodnota proměnné  $Y_1$  chybějící, a  $M = 0$ , když není hodnota proměnné  $Y_1$  chybějící

$U$  je indikátorová proměnná označující, kdy je hodnota proměnné  $Y_2$  větší nebo rovna diskriminační hodnotě  $a$ :  $U = 1$ , když je hodnota proměnné  $Y_2 \geq a$ , a  $U = 0$ , když je hodnota proměnné  $Y_2 < a$ .

**Zdroj: vlastní zpracování podle [9]**

Symbole  $\pi_1$  a  $\pi_2$  označují parametry vyjadřující podmíněné pravděpodobnosti chybějících údajů v proměnné  $Y_1$ . Parametr  $\pi_0$  je pravděpodobnost, že hodnota v proměnné  $Y_2$  je rovná nebo větší než  $a$ , tj.  $P(Y_2 \geq a) = P(U = 1) = \pi_0$ . Pravděpodobnost  $\pi_0$  přímo souvisí s mezním bodem  $a$ . K nastavení hodnoty pro tento parametr  $\pi_0$  je potřebné pouze specifikovat mezní hodnotu  $a$ . Podobně jako u zcela náhodného mechanismu chybění lze pro každý pravděpodobnostní parametr vypočítat rozptyl spojený s odhadovanou hodnotou [9]. Jestliže  $n$  je celkový počet jednotek a  $n_1 = n\pi_0$  je počet jednotek s hodnotami proměnné  $Y_2 \geq a$ , potom příslušné rozptyly pro odhadované  $\pi_0$ ,  $\pi_1$  a  $\pi_2$  jsou:

$$\text{Var}(\Pi_0) = \frac{\pi_0(1-\pi_0)}{n}, \quad (11a)$$

$$\text{Var}(\Pi_1) = \frac{\pi_1(1-\pi_1)}{n} a \quad (11b)$$

$$\text{Var}(\Pi_2) = \frac{\pi_2(1-\pi_2)}{n}. \quad (11c)$$

K určení očekávaného procenta chybějících hodnot v proměnné  $Y_1$ , je nutné nejprve vypočítat nepodmíněnou pravděpodobnost chybějících údajů v proměnné  $Y_1$ :

$$\begin{aligned} \pi_{miss} &= P(M = 1) \\ &= P(M = 1|U = 1)P(U = 1) + P(M = 1|U = 0)(1 - P(U = 1)) \\ &= \pi_1\pi_0 + \pi_2(1 - \pi_0) \end{aligned} \quad (12)$$

Počet zpravodajských jednotek (případů) s chybějícími údaji je náhodná proměnná  $K$  podléhající binomickému rozdělení (předpokládá se nezávislost mezi vznikem chybějících údajů), tj.  $K \sim \text{Bin}(n, \pi_{miss})$ . Očekávané procento chybějících údajů v proměnné  $Y_1$  ve výběrovém souboru je:

$$E\left(\frac{K}{n}\right) = \pi_1\pi_0 + \pi_2(1 - \pi_0), \quad (13)$$

a rozptyl pro odhadované procento (13) chybějících údajů v proměnné  $Y_1$  je:

$$Var\left(\frac{K}{n}\right) = \frac{\pi_{miss}(1-\pi_{miss})}{n}. \quad (14)$$

Očekávaný počet vzorů chybějících dat pro náhodný mechanismus chybění se podobně jako v (9) určí rovnicí:

$$E(D) = 2 - \pi_{miss}^n - (1 - \pi_{miss})^n, \quad (15)$$

K posouzení vlastností náhodného mechanismu chybění patří také i zjištění síly závislosti mezi chybějícími údaji v proměnné  $Y_1$  na pozorovaných hodnotách proměnné  $Y_2$  představované mírou asociace mezi indikátorem chybějících údajů  $M$  a indikátorem  $U$  hodnoty proměnné  $Y_2$ . Pokud bude tato síla závislosti nulová, pozorované hodnoty proměnné  $Y_2$  nemají vliv na chybějící hodnoty proměnné  $Y_1$  a z náhodného mechanismu chybění MAR se stává zcela náhodný mechanismus chybění MCAR (viz obrázek č. 1).

Podobně jako u ostatních binárních proměnných, sílu závislosti mezi indikátorem chybějících údajů  $M$  a indikátorem  $U$  hodnoty proměnné  $Y_2$  lze měřit pomocí absolutního rozdílu pravděpodobností ( $ARD$ ) anebo použitím poměru šancí ( $OR$ ), tj.

$$ARD = \pi_1 - \pi_2 \quad (16)$$

$$OR = \frac{P(M = 1|U = 1)/(1 - P(M = 1|U = 1))}{P(M = 1|U = 0)/(1 - P(M = 1|U = 0))} = \frac{\pi_1/(1 - \pi_1)}{\pi_2/(1 - \pi_2)} \quad (17)$$

Symbol  $ARD$  [1] pro absolutní rozdíl pravděpodobností vychází z anglického názvu *absolute risk difference* a  $OR$  (z anglického názvu *Odds Ratio*) znamená poměr šancí. Velká hodnota  $ARD$ , resp. hodnota  $OR$  větší než 1 (čím dále od 1, tím více) indikuje silnou závislost; je-li  $1 - \pi_1 = 0$  anebo  $1 - \pi_2 = 0$  poměr šancí  $OR$  není definován. Proto k měření síly závislosti se spíše používá ukazatel  $ARD$ .

Rovnice (16) a (17) měří sílu závislosti mezi indikátory (binárními proměnnými)  $M$  a  $U$  na úrovni populace. Na úrovni výběrového souboru se odhadovaná síla závislosti může lišit vzorek od vzorku. Odchytky spojené s odhadovanou  $ARD$  a odhadovaným logaritmem poměru šancí  $OR$  jsou následující (odvození viz [14]):

$$Var(\Pi_1 - \Pi_2) = Var(\Pi_1) + Var(\Pi_2) = \frac{\pi_1(1-\pi_1)}{n} + \frac{\pi_2(1-\pi_2)}{n}, \quad (18)$$

$$Var(\log(OR)) = \frac{1}{\pi_1\pi_0} + \frac{1}{(1-\pi_1)\pi_0} + \frac{1}{\pi_2(1-\pi_0)} + \frac{1}{(1-\pi_2)(1-\pi_0)} \quad (19)$$

kde výrazy ve jmenovateli jsou definovány v kontingenční tabulce na obrázku č. 3.

Jelikož poměr šancí lze vyjádřit také pomocí logistického regresního modelu [1], vztah mezi indikátorem chybějících údajů  $M$  a  $U$  indikátorem hodnoty proměnné  $Y_2$  je

možné také vyjádřit na základě logistické regrese. Logaritmus poměru šancí může  $OR$  být predikován pomocí indikátoru  $U$  takto:

$$\log \left( \frac{P(M=1)}{1-P(M=1)} \right) = \beta_0 + \beta_1 U, \quad (20)$$

kde koeficient  $\beta_0$  je hodnota logaritmu<sup>3</sup> poměru šancí  $OR$  za platnosti  $U = 0$ :

$$\beta_0 = \log \left( \frac{P(M=1/U=0)}{1-P(M=1)/U=0} \right) = \log \left( \frac{\pi_2}{1-\pi_2} \right) \text{ a} \quad (21)$$

kde koeficient  $\beta_1$  je hodnota logaritmu poměru šancí  $OR$ :

$$\beta_1 = \log (OR) = \log \left( \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} \right) \quad (22)$$

Z rovnice (20) vyplývá, že čím je vyšší hodnota koeficientu  $\beta_1$ , tím je silnější závislost mezi indikátory  $M$  a  $U$ . Rovnice (20) také dokazuje, že pravděpodobnostní pravidlo s jednou diskriminační hodnotou je současně ekvivalentní logistickému regresnímu modelu, který může popisovat náhodný mechanismus chybění MAR pomocí logistické regrese.

### Náhodný mechanismus chybění MAR s vícenásobnou rozhodovací hodnotou

U náhodného mechanismu chybění MAR s vícenásobnou rozhodovací hodnotou je třeba zadat více hraničních bodů pro prediktor chybějících dat. Jednou z výhod použití vícenásobné rozhodovací hodnoty je, že model lze použít k vytvoření lineárního i nelineárního vztahu mezi indikátorem chybějících dat  $M$  a prediktorem chybějících dat  $Y_2$ . K nelineárnímu vztahu dochází, když subjekty s extrémními hodnotami na prediktoru chybějících dat mají vyšší nebo nižší pravděpodobnost, že nebudou chybět, než subjekty s hodnotami středního rozsahu na prediktoru. Naproti tomu k lineárnímu vztahu dochází, když se pravděpodobnost, že bude chybět, postupně zvyšuje nebo snižuje s tím, jak se zvyšuje hodnota prediktoru chybějících dat.

Pro vytvoření nelineárního vztahu předpokládejme, že pravděpodobnost chybějících hodnot u proměnné  $Y_1$  závisí na dvou hraničních bodech,  $a$  a  $-a$ , hodnot proměnné  $Y_2$ . Podobně jako u modelu s jednoduchou rozhodovací hodnotou nechť  $M$  je indikátor chybějících dat a  $U$  je indikátor označující, zda je hodnota  $Y_2$  mezi dvěma diskriminačními body, tj.  $U = 1$ , když  $Y_2 \geq a$  anebo  $Y_2 \leq -a$  a  $U = 0$ , když  $-a < Y_2 < a$ . Toto pravidlo chybějících dat je stejné jako pravidlo pro model s jedinou rozhodovací hodnotou. Jinými slovy, v případě nelineárního vztahu může být na chybějící datové pravidlo spojené s vícenásobnými rozhodovacími hodnotami nahlíženo tak, jako na pravidlo chybějících dat spojené s jednoduchou rozhodovací hodnotou. Důsledkem je, že v tomto případě lze všechny rovnice pro model s jednoduchou rozhodovací hodnotou použít i pro model s vícenásobnými rozhodovacími hodnotami [14].

K vytvoření lineárního vztahu mezi indikátorem chybějících dat  $M$  a prediktorem chybějících dat  $Y_2$  je nutné specifikovat alespoň dva hraniční body v prediktoru chybějících dat  $Y_2$ . Ve většině případů se specifikují tři nebo čtyři hraniční body, obvykle kvartilové nebo kvantilové hodnoty prediktoru chybějících dat  $Y_2$ . Proměnná

<sup>3</sup> Symbol  $\log$  označuje logaritmus o základu  $e$ , tj. přirozený logaritmus.

$Y_2$  se rozdělí do čtyř nebo pěti skupin, přičemž hodnoty prediktoru chybějících dat  $Y_2$  u každé zpravodajské jednotky mají stejnou šanci být v kterékoli ze skupin. Přejíždí-li se ze skupiny s nejnižšími hodnotami prediktoru  $Y_2$  do skupiny s nejvyššími hodnotami, pravděpodobnost chybění obvykle roste nebo klesá konstantní rychlostí.

Nechť  $M$  je indikátor chybějících dat a  $V$  je diskretní uniformní náhodná proměnná vytvořená na základě hodnot  $Y_2$  (tj.  $V$  lze považovat za prediktor chybějících dat), potom náhodná proměnná  $V$  vystupující jako prediktor chybějících dat  $Y_2$  nabývá následujících hodnot:

$$V = \begin{cases} 1 & \text{if } Y_2 < Q_1 \\ 2 & \text{if } Q_1 \leq Y_2 < Q_2 \\ 3 & \text{if } Q_2 \leq Y_2 < Q_3 \\ 4 & \text{if } Y_2 \geq Q_3 \end{cases} \quad (23)$$

kde symboly  $Q_1$ ,  $Q_2$ , a  $Q_3$  jsou kvartilové hodnoty prediktoru chybějících hodnot  $Y_2$ . Pravděpodobnostní pravidlo chybějících dat pro proměnnou  $Y_1$  je potom:

$$\begin{aligned} P(M = 1|V = 1) &= \pi_1 \\ P(M = 1|V = 2) &= \pi_2 \\ P(M = 1|V = 3) &= \pi_3 \\ & a \\ P(M = 1|V = 4) &= \pi_4 \end{aligned} \quad (24)$$

Pravděpodobnostní pravidlo chybějících údajů s více diskriminačními hodnotami lze opět nejlépe ilustrovat pomocí kontingenční tabulky pro dva indikátory. Tato kontingenční tabulka je uvedena na obrázku č. 4.

**Obrázek č. 4: Kontingenční tabulka pro náhodný mechanismus neúplných dat s více diskriminačními hodnotami**

V	M	
	1	0
1	$P(M = 1 V = 1)P(V = 1) = \pi_1\pi_0$	$P(M = 0 V = 1)P(V = 1) = (1 - \pi_1)\pi_0$
2	$P(M = 1 V = 2)P(V = 2) = \pi_2\pi_0$	$P(M = 0 V = 2)P(V = 2) = (1 - \pi_2)\pi_0$
3	$P(M = 1 V = 3)P(V = 3) = \pi_3\pi_0$	$P(M = 0 V = 3)P(V = 3) = (1 - \pi_3)\pi_0$
4	$P(M = 1 V = 4)P(V = 4) = \pi_4\pi_0$	$P(M = 0 V = 4)P(V = 4) = (1 - \pi_4)\pi_0$

$M$  je indikátorová proměnná označující, kdy je hodnota proměnné  $Y_1$  chybějící:  $M = 1$ , když je hodnota proměnné  $Y_1$  chybějící, a  $M = 0$ , když není hodnota proměnné  $Y_1$  chybějící

$V$  je diskretní indikátorová proměnná rovnoměrně rozdělená označující, ve kterém kvartilu hodnota proměnné  $Y_2$  je:  $V = 1$  když  $Y_2 < Q_1$ ,  $V = 2$  když  $Q_1 \leq Y_2 < Q_2$ ,  $V = 3$  když  $Q_2 \leq Y_2 < Q_3$  a  $V = 4$  když  $Y_2 \geq Q_3$ , kde  $Q_1$ ,  $Q_2$  a  $Q_3$  jsou kvartilové hodnoty proměnné  $Y_2$ .

**Zdroj: vlastní zpracování podle [9]**

V pravděpodobnostním pravidlu pro chybějící data vystupuje 5 parametrů. Z nich jsou 4 pravděpodobnostní parametry  $\pi_1$ ,  $\pi_2$ ,  $\pi_3$  and  $\pi_4$ . Poslední parametr označený jako  $\pi_0$  představuje pravděpodobnost indikátoru  $V$ , tj.  $P(V = i) = \pi_0 = 0,25$ , kde  $i \in \{1,2,3,4\}$ . Hodnota parametru  $\pi_0$  je rovna 0,25, protože prediktor chybějících hodnot proměnná  $Y_2$  je kvartilovými mezními hodnotami rozdělena na 4 stejné intervaly. Pro každý parametr je možné odhadnout rozptyl asociovaný s odhadem počtu procent chybějících údajů. Označí-li se  $n$  jako celkový počet hodnot v prediktoru chybějících hodnot, tj. v proměnné  $Y_2$ , potom počet hodnot v každém kvartilovém intervalu je  $n_0 = \pi_0 n = 0,25n$ . Rozptyl odhadovaného parametru  $\pi_0$  je definován rovnicí:



$$Var(\Pi_0) = \frac{\pi_0(1-\pi_0)}{n}, \quad (25)$$

Rozptyly odhadů parametrů  $\pi_j$  pro  $j \in \{1,2,3,4\}$  jsou dány jako:

$$Var(\Pi_j) = \frac{\pi_j(1-\pi_j)}{n}, \quad (26)$$

Odhad pravděpodobnosti chybění hodnoty u každé zpravodajské jednotky lze provést pomocí marginální pravděpodobnosti, že indikátor chybění bude roven 1, tj.:

$$\begin{aligned} \pi_{miss} &= P(\mathbf{M} = 1) \\ &= P(\mathbf{M} = 1|\mathbf{V} = 1)P(\mathbf{V} = 1) + P(\mathbf{M} = 1|\mathbf{V} = 2)P(\mathbf{V} = 2) \\ &\quad + P(\mathbf{M} = 1|\mathbf{V} = 3)P(\mathbf{V} = 3) + P(\mathbf{M} = 1|\mathbf{V} = 4)P(\mathbf{V} = 4) \\ &= \pi_1\pi_0 + \pi_2\pi_0 + \pi_3\pi_0 + \pi_4\pi_0 \end{aligned} \quad (27)$$

Počet zpravodajských jednotek (případů) s chybějícími údaji v proměnné  $Y_1$  je náhodná proměnná  $K$  podléhající binomickému rozdělení (předpokládá se nezávislost vzniků chybějících údajů), tj.  $K \sim Bin(n, \pi_{miss})$ , kde  $n$  je celkový počet hodnot. Očekávané procento chybějících údajů v proměnné  $Y_1$  ve výběrovém souboru je:

$$E\left(\frac{K}{n}\right) = \pi_1\pi_0 + \pi_2\pi_0 + \pi_3\pi_0 + \pi_4\pi_0, \quad (28)$$

a rozptyl pro odhadované procento (28) chybějících údajů v proměnné  $Y_1$  je:

$$Var\left(\frac{K}{n}\right) = \frac{\pi_{miss}(1-\pi_{miss})}{n}. \quad (29)$$

Očekávaný počet jedinečných vzorů chybějících dat pro náhodný mechanismus chybění MAR s kvantilovými hodnotami se podobně jako v (9) určí rovnicí:

$$E(D) = 2 - \pi_{miss}^n - (1 - \pi_{miss})^n, \quad (30)$$

Sílu závislosti mezi indikátorem chybějících údajů  $\mathbf{M}$  a indikátorem  $\mathbf{V}$  hodnoty prediktoru proměnné  $Y_2$  lze měřit pomocí průměrného absolutního rozdílu pravděpodobností *AARD* anebo použitím poměru šancí *OR*. Pro pravděpodobnostní pravidlo využívající kvartilové hodnoty jako hodnoty rozhodovací je *AARD* definována ve tvaru:

$$AARD = \frac{\pi_4 - \pi_1}{3}, \quad (31)$$

V poměru šancí *OR* vystupují marginální pravděpodobnosti pro indikátor chybění  $\mathbf{M}$  podle (27) jako pravděpodobnosti  $\pi_{miss} = P(\mathbf{M} = 1)$ , že údaj bude chybět, tj.:

$$\log(OR) = \log\left(\frac{P(\mathbf{M}=1)}{1-P(\mathbf{M}=1)}\right) \quad (32)$$

Náhodný mechanismus chybění s vícenásobnými diskriminačními hodnotami je přímým rozšířením modelu s jednoduchou rozhodovací hodnotou.

### Náhodný mechanismus chybění MAR s rozhodovacími percentily

Použití percentilů k rozhodování je rozšířením modelu s vícenásobnou rozhodovací hodnotou. V percentilové metodě závisí pravděpodobnost, že u zpravodajské jednotky bude hodnota chybět, na percentilovém pořadí v prediktoru chybějících dat  $Y_2$ . Proto na percentilovou metodu lze pohlížet jako na metodu s vícenásobnými rozhodovacími hodnotami, kde každá zpravodajská jednotka má svůj vlastní hraniční bod na základě percentilového pořadí hodnot prediktorové proměnné.

Nechť pravděpodobnost, že zpravodajská jednotka bude mít chybějící hodnotu v proměnné  $Y_1$ , bude závislá na percentilovém pořadí hodnot prediktoru chybějících dat, tj. proměnné  $Y_2$ . Nechť pravděpodobnost, že zpravodajská jednotka bude mít chybějící hodnotu v proměnné  $Y_1$  bude závislá na percentilovém pořadí hodnot prediktoru chybějících dat, tj. proměnné  $Y_2$ . Nechť náhodná proměnná  $M$  bude v úloze indikátoru chybějících dat. Bude-li přímý vztah mezi indikátorem chybějících údajů a prediktorem chybějících údajů  $Y_2$ , potom pravděpodobnostní pravidlo pro chybějící údaj bude: jestliže hodnota v závislé proměnné bude odpovídat  $k$ -tému percentilu proměnné  $Y_2$ , potom pravděpodobnost, že údaj bude nevyplněn v proměnné  $Y_1$ , odpovídá  $k\%$  pravděpodobnosti chybění údaje v proměnné  $Y_1$ , tj.  $P(M = 1|Y_2 = q_k) = k/100$ , kde  $q_k$  je hodnota proměnné  $Y_2$  korespondující  $k$ -tému percentilu. Bude-li nepřímý vztah mezi indikátorem chybějících údajů a prediktorem chybějících údajů, potom pravděpodobnostní pravidlo pro chybějící údaj bude: jestliže hodnota v závislé proměnné bude odpovídat  $k$ -tému percentilu proměnné  $Y_2$ , potom pravděpodobnost, že údaj bude nevyplněn v proměnné  $Y_1$ , odpovídá  $(100 - k)\%$  pravděpodobnosti chybění údaje v proměnné  $Y_1$ , tj.  $P(M = 1|Y_2 = q_k) = 1 - k/100$ , kde  $q_k$  je hodnota proměnné  $Y_2$  korespondující  $k$ -tému percentilu. Uvedená pravděpodobnostní pravidla chybění jsou jedinými možnými pravděpodobnostními pravidly spojenými s percentilovými hodnotami. Jiné parametry než výše uvedené, se při aplikaci náhodného mechanismu chybění s percentilovými hodnotami nepoužívají.

Aby bylo možné vypočítat pravděpodobnost chybějících dat, je potřebné stanovit pravděpodobnostní rozdělení percentilové řady hodnot proměnné  $Y_2$ . K tomu lze využít empirickou distribuční funkci  $F(\cdot)$  proměnné  $Y_2$ , která mapuje hodnoty proměnné  $Y_2$  do řady percentilů. Percentilové hodnoty proměnné  $Y_2$  proto mají standardní rovnoměrné rozdělení, tj.:

$$F(Y_2) \sim Unif(0,1). \quad (33)$$

Toho důsledkem je, že očekávaná percentilové pořadí zpravodajské jednotky je 50. percentil, a tedy pravděpodobnost chybějících dat je vždy 50 %, tj.  $P(M = 1) = 0,5$ . Podobně jako u předešlých modelů počet zpravodajských jednotek (případů) s chybějícími údaji v proměnné  $Y_1$  je náhodná proměnná  $K$  podléhající binomickému rozdělení (předpokládá se nezávislost mezi vznikem chybějících údajů), tj.  $K \sim Bin(n, 0,5)$ , kde  $n$  je celkový počet hodnot. Očekávané procento chybějících údajů v proměnné  $Y_1$  ve výběrovém souboru j:

$$E\left(\frac{K}{n}\right) = 0,5, \quad (34)$$

a rozptyl pro odhadované procento (34) chybějících údajů v proměnné  $Y_1$  je:

$$Var\left(\frac{K}{n}\right) = \frac{0,5(1-0,5)}{n} = \frac{0,25}{n}. \quad (35)$$

Očekávaný počet jedinečných vzorů chybějících dat pro náhodný mechanismus chybění MAR s percentilovými hodnotami se podobně jako v (9) určí rovnicí

$$E(D) = 2 - \pi_{miss}^n - (1 - \pi_{miss})^n = 2 - 0,5^n - (1 - 0,5)^n. \quad (36)$$

U percentilové metody se nedá měnit síla závislosti mezi indikátorem chybějících dat  $M$  a prediktorovou proměnou  $Y_2$ . Způsobem, jak kvantifikovat sílu závislosti mezi indikátorem chybějících dat a proměnnou  $Y_2$  jako prediktorovou proměnnou vytvořené percentilovými hodnotami, je najít model logistické regrese, který při dostatečně velkém  $n$  aproximuje pravděpodobnostní pravidlo chybějících dat [14], tj.:

$$\log \left( \frac{P(M=1)}{1-P(M=1)} \right) = 1,70 \cdot Y_2 \quad (37)$$

### Náhodný mechanismus chybění MAR jako model logistické regrese

Jak je patrné z rovnic (22), (32) a (37), pravděpodobnostní pravidla chybějících údajů uvedená v předešlých podkapitolách se dají přeformulovat do tvaru modelů logistické regrese. Modeluje-li se náhodný mechanismus chybění přímo pomocí logistické regrese, lze považovat logistický regresní model přímo jako pravděpodobnostní pravidlo chybějících údajů a populační regresní koeficienty asociované s modelem jako parametry tohoto pravděpodobnostního pravidla.

Jestliže pravděpodobnost chybějící hodnoty v proměnné  $Y_1$   $i$ -té zpravodajské jednotky je vztažena k prediktoru chybějících hodnot (proměnné  $Y_2$ ), potom regresní logistický model pro  $i$ -tou zpravodajskou jednotku je definován jako:

$$\log \left( \frac{P(M_i = 1|y_{2,i})}{1-P(M_i = 1|y_{2,i})} \right) = \beta_0 + \beta_1 \cdot y_{2,i}, \quad (38)$$

kde  $M_i$  je indikátor chybějící hodnoty a  $y_{2,i}$  je hodnota proměnné  $Y_2$  u  $i$ -té zpravodajské jednotky. Parametry asociované s pravděpodobnostním pravidlem chybějících údajů jsou  $\beta_0$  a  $\beta_1$ . V závislosti na hodnotě proměnné  $Y_2$  je pravděpodobnost, že údaj v proměnné  $Y_2$  u  $i$ -té zpravodajské jednotky chybí, definována rovnicí:

$$P(M_i = 1|y_{2,i}) = \frac{1}{1+e^{-\beta_0-\beta_1 y_{2,i}}} \quad (39)$$

Pro velké velikosti výběrového souboru (cca 10 000 zpravodajských jednotek) lze odhadnout očekávané procento chybějících hodnot průměrem pravděpodobností ve tvaru:

$$\pi_{miss} = \frac{1}{n} \sum_{i=1}^n P(M_i = 1|y_{2,i}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{1+e^{-\beta_0-\beta_1 y_{2,i}}}. \quad (40)$$

Z hlediska síly závislosti mezi indikátorem chybějících údajů  $M$  a prediktorovou proměnnou  $Y_2$  vyšší hodnoty  $\beta_1$  indikují silnější závislost. Logistický regresní model však nelze odhadnout, když  $P(M_i = 1|y_{2,i}) = 1$  nebo  $P(M_i = 0|y_{2,i}) = 1$ , protože logaritmus šance indikátoru  $M$  není v těchto případech definován nebo se rovná nekonečnu. Výhodou logistické regrese je, že pravděpodobnost chybějících hodnot se postupně mění s tím, jak se mění hodnota prediktoru chybějících dat, což vytváří

realističtější situaci vzhledem k metodám popsaným v předešlých podkapitolách. Nevýhodou metody logistické regrese je však to, že neumožňuje nastavit příliš silnou závislost mezi indikátorem chybějících dat  $M$  a prediktorovou proměnnou  $Y_2$ .

### Statistické modely nenáhodného mechanismu chybění (NMAR)

Nenáhodný mechanismus chybějících dat (NMAR) se vyznačuje tím, že výskyt chybějících hodnot závisí na hodnotách  $y_{miss}$ , které nejsou pozorované. Nenáhodný mechanismus chybění dat představuje podmíněné rozdělení pravděpodobnosti indikátorové proměnné  $M$  vzhledem k pozorované proměnné  $Y$  určené neznámými parametry  $\varphi$ , tj.  $f(m|y_{obs}, y_{miss}; \varphi)$ .

Modelovat mechanismus chybějících hodnot umožňují dva rozdílné přístupy určené právě pro data, která nechybí náhodně.

- **První přístup představují tzv. modely výběru (selection models).** U těchto modelů se v prvním kroku specifikuje rozdělení potenciálně kompletních dat (tzn. datové matice  $Y$ , která je složená z pozorovaných a chybějících hodnot). V dalším kroku se podle [11] specifikuje model [4], podle kterého závisí výskyt chybějících hodnot na datové matici  $Y$ .
- **Druhý přístup tvoří tzv. modely smíšených vzorů (pattern-mixture models).** U tohoto přístupu jsou jednotky rozdělené do skupin podle jedinečných vzorů chybějících hodnot. Následně se v každé ze skupin provádí statistická analýza [2]. Použitý pojem „smíšené vzory“ má indikovat, že výsledné marginální rozdělení datové matice  $Y$  je směsí několika rozdělení [4].

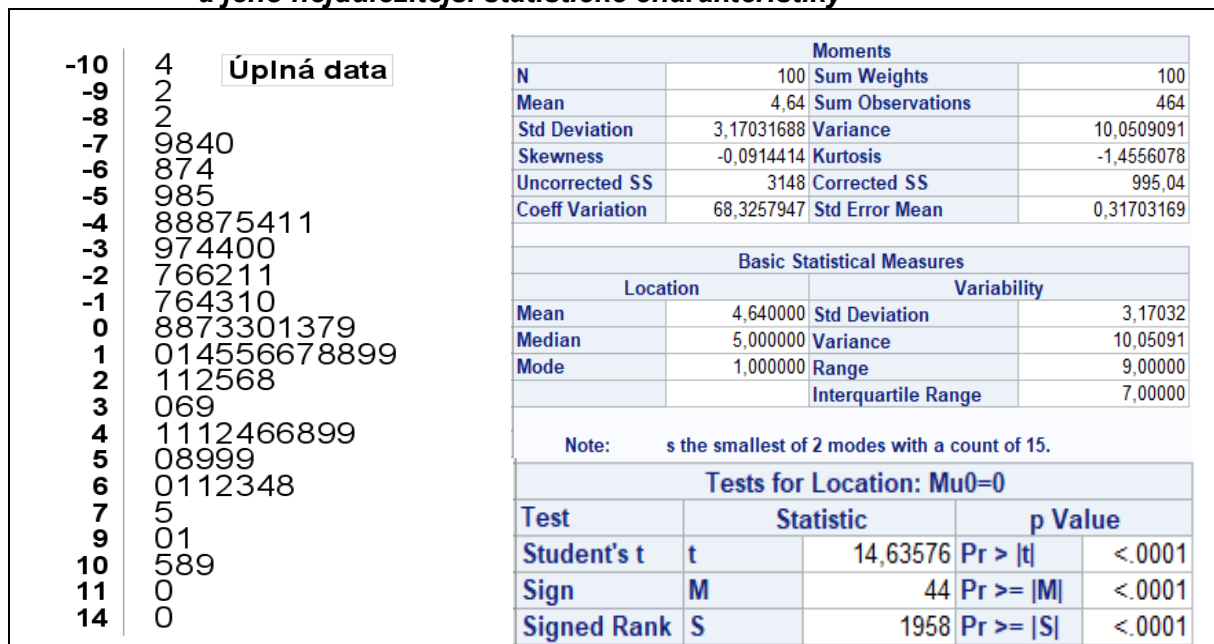
Podrobnější analýza nenáhodného mechanismu chybění v neúplných datech přesahuje rozsah tohoto článku. Detailnější informace o mechanismu NMAR lze dohledat například v [5] nebo v [2].

## 7. UKÁZKA STATISTICKÝCH MODELŮ CHYBĚJÍCÍCH DAT

### Data použitá k ukázce statistických modelů chybějících údajů

Pro účely statistického modelování chybějících údajů byla vygenerována umělá data. Jedná se o datový vzorek se 100 hodnotami v rozsahu  $-10,4$  až  $14,1$  u proměnné  $Y_1$ , pro které jsou jednotlivé modely chybění dat modelovány, a v rozsahu 1 až 100 u prediktorové proměnné  $Y_2$ . Data jsou generována z normálního rozdělení  $N(0, 5)$ . Rozdělení pravděpodobnosti  $\pi_{miss}$  chybějících hodnot v proměnné  $Y_1$  je modelováno pomocí generátorů náhodné proměnné s rozdělením pravděpodobnosti pro jednotlivé typy mechanismu chybění. V rámci modelování mechanismu náhodného chybění MAR bude pro prediktorovou proměnnou  $Y_2$  použito pravděpodobnostní pravidlo s rovnoměrným rozdělením pravděpodobnosti. Pro hodnocení statistických modelů se budou posuzovat zejména momentové charakteristiky a statistické míry sloužící pro měření závislosti náhodných proměnných. Statistické modely chybějících dat a jejich měření (včetně statistických grafů) bylo realizováno pomocí funkcí a procedur statistického systému SAS. Nejdůležitější charakteristiky generovaného vzorku dat pro statistickou analýzu chybějících údajů ilustruje obrázek č. 5.

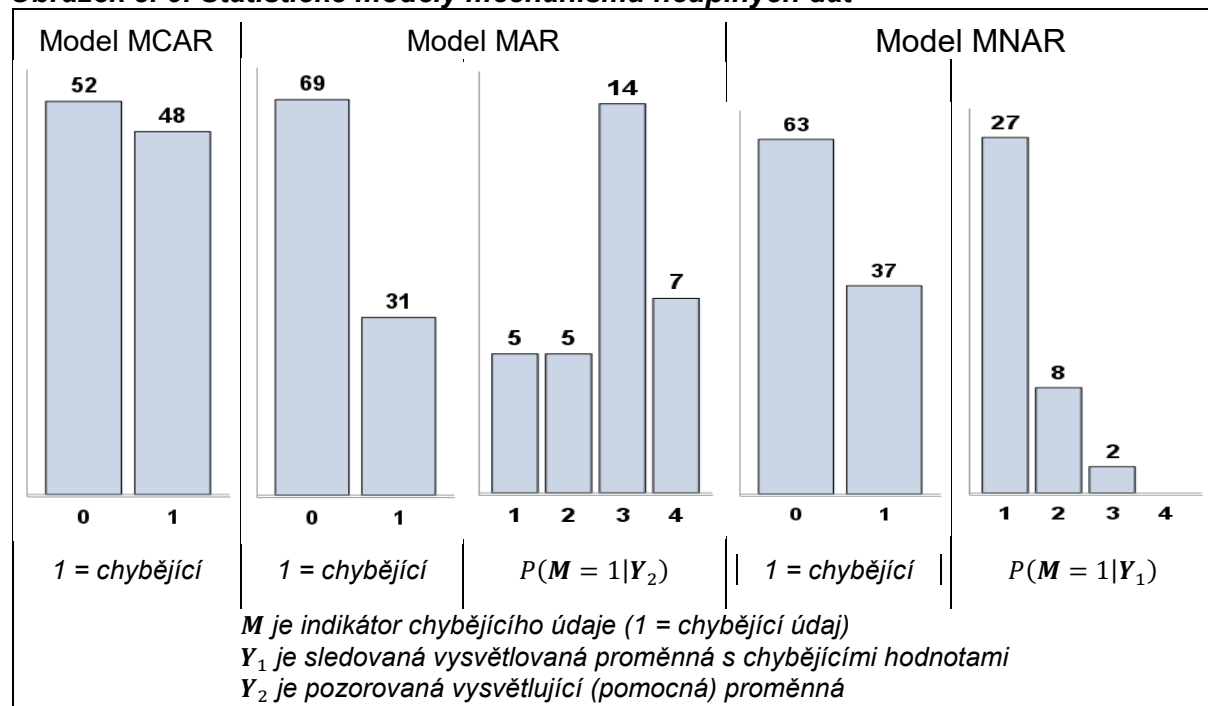
**Obrázek č. 5: Rozdělení generovaných dat (stem-and-leaf graf) bez chybějících údajů a jeho nejdůležitější statistické charakteristiky**



Zdroj: vlastní zpracování podle [6]

Na levé straně obrázku č. 5 je tzv. stem-and-leaf graf<sup>4</sup>, který vizualizuje rozdělení generovaných hodnot. Na pravé straně obrázku č. 5 jsou zaznamenány nejdůležitější charakteristiky vzorku generovaných dat.

**Obrázek č. 6: Statistické modely mechanismů neúplných dat**



Zdroj: vlastní zpracování podle [6]

<sup>4</sup> Stem-and-leaf graf je statistický graf podobný histogramu, data jsou uspořádána do tzv. stonků (angl. stems) a listů (angl. leaves). Stonky mohou například být celé části reálných čísel a stonky obsahují číslice, které patří do desetinného rozvoje odpovídajících reálných čísel.

Princip činnosti jednotlivých modelů neúplných dat reprezentuje obrázek č. 6. Zcela náhodný model neúplných dat (označený MCAR) vykázal ze 100 generovaných hodnot 48 hodnot chybějících a 52 pozorovaných hodnot. U tohoto modelu chybějící hodnoty jsou zcela nezávislé jak od ostatních zjišťovaných proměnných, tak i od rozdělení hodnot v samotné sledované proměnné s chybějícími daty. Jedná se proto o tzv. ignorovatelný mechanismus chybění, který má nejmenší vliv na odhadované hodnoty. U modelu neúplných dat MAR byl zjištěn menší počet chybějících hodnot. Ze 100 generovaných hodnot bylo 69 pozorovaných a 31 chybějících (indikátor chybějících hodnot v grafu má hodnotu 1). U modelu MAR jsou chybějící hodnoty závislé na prediktorové proměnné  $Y_2$ , na jejímž rozdělení výskyt chyb záleží. V případě použitého statistického modelu chybění MAR jako prediktorová proměnná byla použita proměnná  $Y_2$  s rovnoměrným rozdělením pravděpodobnosti. Zjištěné neúplné hodnoty u náhodného modelu chybění MAR byly rozděleny v souladu s předpokládaným rozdělením prediktorové (pomocné) proměnné  $Y_2$ . Toto je patrné na obrázku č. 6 v pravém grafu statistického modelu neúplnosti dat MAR, kdy prediktorová proměnná  $Y_2$  byla rozdělena na 4 stejné intervaly (označené indikátory o hodnotách 1, 2, 3 a 4), každý se stejnou pravděpodobností vzniku chybějící hodnoty (ve skutečnosti v prvním intervalu hodnot proměnné  $Y_2$  se objevilo 5 chybějících hodnot proměnné  $Y_1$ , ve druhém také 5, ve třetím intervalu 14 a v intervalu nejvyšších hodnot prediktorové proměnné  $Y_2$  bylo 7 chybějících hodnot  $Y_2$ ). Mechanismus chybění MAR patří také k tzv. ignorovatelným mechanismům, které lze při odhadech parametrů zanedbat. Při odhadech parametrů s tímto typem chybějících dat je však nutné brát do úvahy i význam prediktorové (pomocné) proměnné, na jejímž rozdělení výskyt chybějících údajů také záleží, tj. rozdělení chybějících dat je podmíněno rozdělením prediktorové (pomocné) proměnné. Nenáhodný mechanismus chybění (na obrázku č. 6 označený jako MNAR) již při statistické inferenci zanedbat nelze. Nenáhodný mechanismus chybění MNAR z obrázku č. 6 vykazuje 37 chybějících hodnot ve sledované proměnné  $Y_1$  v generovaném vzorku (viz pravý graf modelu MNAR). Uvedených 37 chybějících hodnot bylo rozdělených v intervalech nižších hodnot (v grafu označených jako interval 1, 2 a 3) sledované proměnné  $Y_1$  (v prvním intervalu hodnot bylo 27 chybějících hodnot, ve druhém intervalu bylo zaznamenáno 8 chybějících hodnot a ve třetím chyběly 2 hodnoty). V praxi je podle [12] mnohdy složité určit rozdělení proměnné  $Y_1$ , jejíž některé hodnoty jsou chybějící. Chybějící data ve sledované proměnné  $Y_1$  závisí na neznámém rozdělení samotné sledované proměnné  $Y_1$ , která je předmětem statistické inference (viz pravý graf modelu MNAR). Odhady z takto pořízených dat jsou proto nejvíce zatíženy chybami.

**Obrázek č. 7: Momentové charakteristiky dat s různými mechanismy chybění**

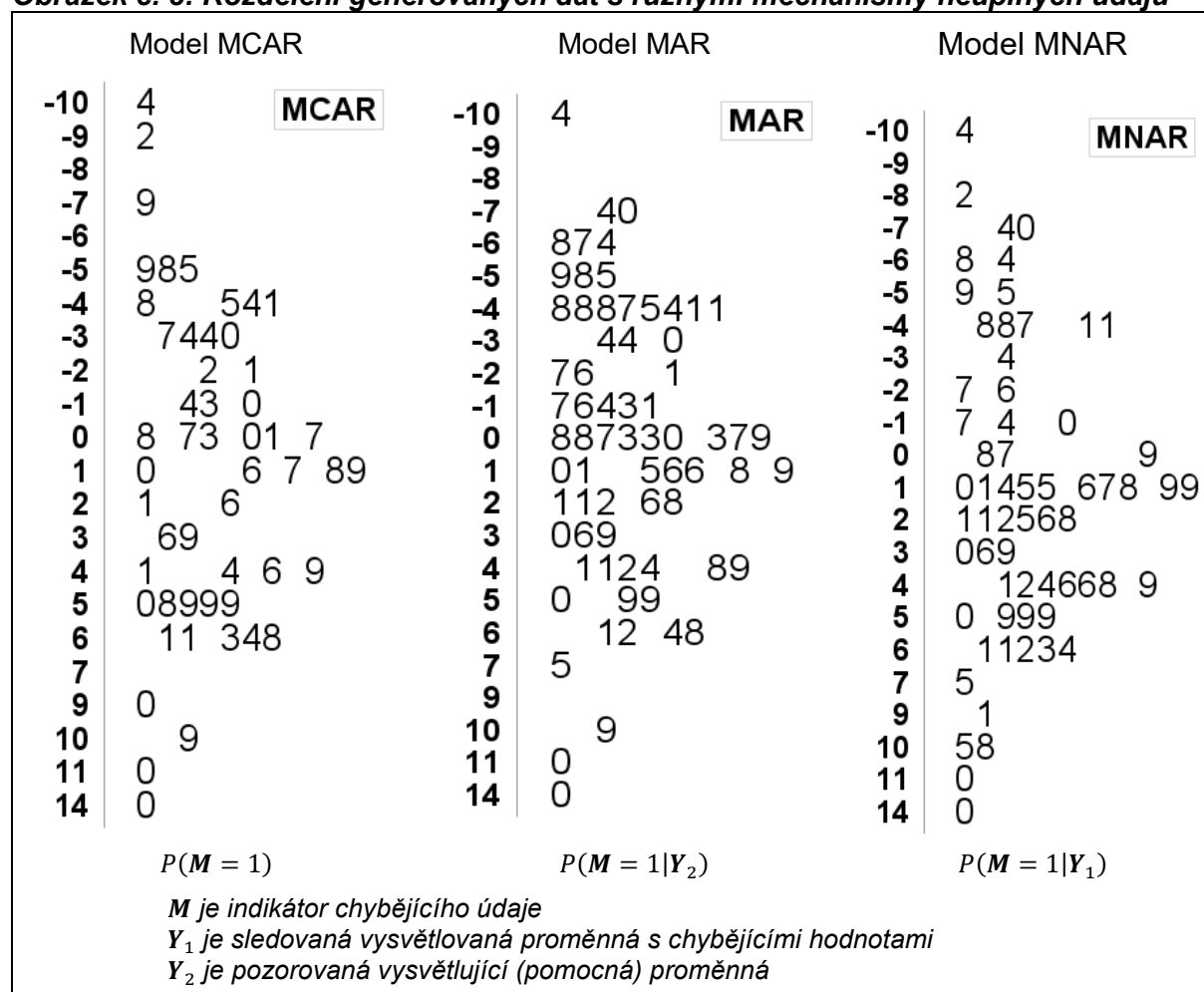
Moments	Mechanismus chybění			
	Úplná data	MCAR	MAR	MNAR
N	100	52	69	63
Mean	4,640	4,596	4,666	4,730
Std Deviation	3,170	3,309	3,118	3,122
Skewness	-0,091	-0,044	-0,117	-0,134
Uncorrected SS	3 148	1 657	2 164	2 014
Coeff Variation	68,326	72,001	66,826	66,007
Sum Weights	100	52	69	63
Sum Observations	464	239	322	298
Variance	10,051	10,951	9,725	9,784
Kurtosis	-1,456	-1,494	-1,411	-1,368
Corrected SS	995,04	558,52	661,33	604,41
Std Error Mean	0,317	0,459	0,375	0,393

**Zdroj: vlastní zpracování podle [6]**

Možné odchylky od skutečných hodnot odhadovaných parametrů ilustruje obrázek č. 7. Největší odchylkou od průměrné hodnoty zjištěné z úplných dat (bez chybějících údajů) vykazují data, v nichž byl modelován nenáhodný mechanismus chybění MNAR. Výběrový průměr úplných dat byl zjištěn ve výši 4,640, výběrový průměr dat s mechanismem chybění MNAR byl ve výši 4,730. Nejvyšší variabilita odhadů byla zjištěna pro data se zcela náhodným mechanismem chybění MCAR. Z hlediska statistické inference nad daty s chybějícími údaji vedlo použití znalosti o rozdělení prediktorové (pomocné) proměnné k nejlepším výsledkům, a to jak v přesnosti, tak i variabilitě odhadů. K průměru z úplných dat se nejvíce blíží výběrový průměr z dat s neúplností typu MAR, přičemž variabilita tohoto průměru MAR je nižší než u dat se zcela náhodným mechanismem chybění MCAR.

Jednotlivé mechanismy chybějících údajů v datech mají také vliv na rozdělení hodnot sledované proměnné. Jak působí jednotlivé typy chybových mechanismů na generovaná data, reprezentuje obrázek č. 8 pomocí tzv. stem-and-leaf grafů.

**Obrázek č. 8: Rozdělení generovaných dat s různými mechanismy neúplných údajů**



**Zdroj: vlastní zpracování podle [6]**

Pro reálnou praxi je také důležité zjištění, zda opravdu existuje souvislost mezi pravděpodobností  $\pi_{miss} = P(M = 1)$ , že hodnota ve sledované (neúplné) proměnné  $Y_1$  bude chybět, a prediktorovou (vysvětlující) proměnnou  $Y_2$  v náhodném mechanismu chybění MAR, resp. že hodnota ve sledované (neúplné) proměnné  $Y_1$  bude chybět a současně závisí na neznámém rozdělení samotné sledované proměnné

$Y_1$ , která je předmětem statistické inference, jak tomu je v nenáhodném mechanismu chybění MNAR. Za účelem zjištění míry závislosti lze využít standardní statistické testy, a to  $\chi^2$ -kvadrát test dobré shody, resp. Fisherův exaktní test (při očekávaných hodnotách v buňkách kontingenční tabulky  $< 5$ ) a další míry asociace. Výsledky testování hypotézy o nezávislosti náhodných proměnných a další koeficienty jsou uvedeny na obrázku č. 9. Na základě výsledků uvedených na obrázku č. 9 lze považovat pravděpodobnost, že hodnota ve sledované proměnné bude chybějící, za pravděpodobnost podmíněnou:

- u statistického modelu MAR na rozdělení hodnot prediktorové proměnné  $Y_2$ ,
- u statistického modelu MNAR na neznámém rozdělení samotné sledované proměnné  $Y_1$ .

**Obrázek č. 9: Míry statistické závislosti statistických modelů MAR a MNAR**

Statistic	MAR			MNAR		
	DF	Value	Prob	DF	Value	Prob
Chi-Square	3	10,0808	0,0179	2	11,5802	0,0031
Likelihood Ratio Chi-Square	3	9,6504	0,0218	2	11,9903	0,0025
Mantel-Haenszel Chi-Square	1	1,8462	0,2060	1	8,3577	0,0038
Phi Coefficient		0,3175			0,3403	
Contingency Coefficient		0,3026			0,3222	
Cramer's V		0,3175			0,3403	
<b>WARNING: 70% of the cells have expected counts less than 5. (Asymptotic) Chi-Square may not be a valid test.</b>						
<b>Fisher's Exact Test</b>						
Table Probability (P)		<.0001			0,0002	
Pr <= P		0,0246			0,0022	

**Zdroj: vlastní zpracování podle [6]**

## 8. ZÁVĚR

V současné době jsou neúplné (chybějící) údaje frekventovanou součástí souborů dat a jejich výskytu není přikládán 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). Problematické je také i stanovení směrodatných chyb takto určených odhadů [9].

Metody analýzy neúplných dat používané k zajištění adekvátní statistické inference musí vycházet z konkrétní situace v zaznamenaných datech a jejich uplatnění je závislé na více faktorech. Obecně jsou metody analýzy neúplných dat podle [5] seskupeny do následujících kategorií, jejichž použití se vzájemně nevylučuje:

- **Postupy založené na jednotkách s kompletně vyplněnými odpověďmi:** Zpravodajské jednotky s neúplně vyplněnými údaji se ze statistických analýz vypouští a k dalšímu zpracování se použijí pouze jednotky s kompletně zjištěnými údaji. Tuto strategii, která se nazývá analýzou úplných případů, je snadné implementovat a může být uspokojivá s malým množstvím chybějících údajů. Může však vést k vážným zkreslením a obvykle není příliš efektivní, zejména při odhadování závěrů pro subpopulace.
- **Postupy založené na modifikaci vah pro jednotlivé funkce odhadu:** Do funkcí pro odhad populačních parametrů jsou použity nejen pravděpodobnostní váhy



(pravděpodobnosti zahrnutí jednotek), ale také i váhy odvozené z neodpovědí zpravodajských jednotek.

- **Imputace chybějících hodnot:** Chybějící hodnoty jsou vyplněny a výsledná doplněná data jsou analyzována standardními metodami. Mezi běžně používané postupy pro imputaci patří imputace hot deck, kdy se k imputaci používají jednotky s kompletními údaji; imputace průměrnou hodnotou, které vycházejí z průměrných hodnot kompletně vyplněných jednotek; a regresní imputace, kde se chybějící proměnné pro jednotku odhadují předpokládanými hodnotami z regrese na dostupných proměnných pro tuto jednotku.
- **Metody založené na statistických modelech:** Široká třída metod založených na definici statistického modelu pro úplná data vycházející z funkce věrohodnosti nebo aposteriorního rozdělení. Parametry modelu jsou odhadovány takovými metodami, jako je metoda maximální věrohodnosti anebo expektační-maximalizační algoritmy, apod.

## LITERATURA

- [1] AGRESTI, A. – KATERI, M.: Categorical data analysis. In: International Encyclopedia of Statistical Science. Springer, 2011, s. 206 – 208. ISBN 978-3-642-04897-5.
- [2] ALLISON, P. D.: Missing Data. Thousand Oaks, CA: Sage. Sage University Papers Series. Quantitative Applications in the Social Sciences, 2001, č. 07-136. ISBN 0-7619-1672-5.
- [3] ENDERS, C. K.: Applied Missing Data Analysis, Second Edition. New York: Guilford Press, 2022. 546 s. ISBN 978-1-462-54986-3.
- [4] GRAHAM, J. W.: Missing data: Analysis and design. New York: Springer, 2012. ISBN 978-1-4614-4017-8.
- [5] LITTLE, R. J. A.: Pattern-Mixture Models for Multivariate Incomplete Data. In: Journal of the American Statistical Association, 1993, č. 421, s. 125 – 134.
- [6] LITTLE, R. J. A. – RUBIN, D., B.: Statistical Analysis with Missing Data, 3rd Edition. New York: John Wiley & Sons, Inc. 2019. ISBN 978-0-470-52679-8.
- [7] McKNIGHT, P. E. – McKNIGHT, K. M. – SIDANI, S. – FIGUEREDO, A. J.: Missing data: a gentle introduction. New York: Guilford Press, 2007. 251 s. ISBN 978-1-59385-394-5.
- [8] McLAWHORN, J. – GRIGSBY, T. J.: Missing Data Techniques and the Statistical Conclusion Validity of Survey-Based Alcohol and Drug Use Research Studies: A Review and Comment on Reproducibility. In: Journal of Drug Issues, 2019, č. 1, s. 44 – 56.
- [9] NAKAGAWA, S.: Missing data: Mechanisms, methods, and messages. In: G. A. Fox, S. Negrete-Yankelevich, & V. J. Sosa (Eds.). Ecological statistics: contemporary theory and application, 2015, s. 81 – 105.
- [10] PECÁKOVÁ, I.: Problém chybějících dat v dotazníkových šetřeních. In: Acta Oeconomica Pragensia. 2014, č. 6, s. 66 – 78.
- [11] RUBIN, D. A.: Inference and Missing Data. In: Biometrika, 1976, č. 3, s. 581 – 592.
- [12] SCHAFER, J. L. – GRAHAM, J. W.: Missing data: Our view of the state of the art. In: Psychological Methods, 2002, č. 2, s. 147 – 177.
- [13] van BUUREN, S.: Flexible Imputation of Missing Data, 2nd Edition. New York: CRC Press, 2018. 444 s. ISBN 978-1-138-58831-8.
- [14] ZHANG, X.: How to generate missing data for simulation studies. In: The Quantitative Methods for Psychology, 2023, č. 2, s. 100 – 122.

## 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.

I když mechanismus chybějících hodnot představuje nejdůležitější faktor, který má nejvýraznější vliv na úspěšnost rozmanitých metod práce s chybějícími hodnotami, není mnohdy jednoduché pochopit příčiny vzniku a dynamiku neúplných dat. Datoví analytici zpravidla nikdy nemohou s jistotou znát mechanismus, podle kterého analyzovaná data chybí. V rámci vykonávání odpovídající statistické analýzy dat za podmínek chybějících údajů se vždy vyplatí podrobně promyslet, který z výše uvedených mechanismů chybění je v dané situaci nejrealističtější. Pro adekvátní analýzu dat je dále vhodné hledat ve výběrovém souboru proměnné, které korelují s výskytem nevyplněných hodnot u proměnných vstupujících do procesu analýz. Při realizaci výběrových šetření je užitečné zavést takové postupy a mechanismy, které by alespoň z části vyloučily situace potenciálně vedoucí ke vzniku chybějících hodnot u klíčových proměnných, například včasnými kontrolami sbíraných dat, řešení otázek respondentů ke sběru dat, vhodnou organizací výběrového šetření atd. Je třeba zdůraznit, že právě předcházení vzniku chybějících hodnot může být tím nejlepším řešením problémů plynoucích z chybějících hodnot [13].

## RESUME

Missing data is especially a problem for the analyses using multivariate statistical methods. Although the incomplete data in the random sample may be represented by a relatively small percentage, this situation may lead to a relatively very small complete data set in the surveyed 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. Omission of units from the analyses can have significant negative impacts - reducing the accuracy of estimates and the power of the performed statistical tests - and can lead to biased results inappropriate for generalization to the target population.

Although the mechanism of missing values is the most important factor that affect the most significantly the success of various methods of working with missing values, it is often not easy to understand the causes and dynamics of incomplete data. As a rule, data analysts can never know with certainty the mechanism by which the analysed data is missing. As part of conducting an appropriate statistical analysis of data under conditions of missing data, it is always worth considering in detail which of the above missing data mechanisms is the most realistic in a given situation. For an adequate data analysis, it is also advisable to look for variables in

the sample set that correlate with the occurrence of unfilled values for the variables entering the analysis process. When conducting sample surveys, it is useful to establish such procedures and mechanisms that would at least partially exclude situations potentially leading to the emergence of the missing values for key variables, for example by timely checks of the collected data, solving respondent's questions about data collection, suitable organization of sample surveys, etc. It should be emphasized that the best solution to the issues arising from missing values seems to be the prevention of missing values.

### **PROFESIJNÝ ŽIVOTOPIS**

*Ing. Roman Pavelka, PhD., v rokoch 1995 – 2010 pracoval v poradenskej spoločnosti Trexima, s. r. o. Na pozícii štatistik – analytik sa zaoberal najmä analýzami mzdových a personálnych údajov. Podieľal sa na tvorbe pravidelných štatistických prehľadov a reportov. Spolupracoval s akademickými pracoviskami, agentúrami i súkromnými subjektami na realizácii a vyhodnocovaní ad hoc štatistických výskumov. Oblasť jeho vedeckého záujmu predstavujú výberové zisťovania, odhady a štatistické modely. V rokoch 2012 až 2013 sa zúčastnil zahraničnej stáže vo Veľkej Británii. Od roku 2013 pôsobil v Národnom ústave certifikovaných meraní vzdelávania (NÚCEM), kde zaisťoval štatistické vyhodnocovanie výsledkov testovania žiakov a študentov. Od roku 2015 pracuje v odbore metód štatistických zisťovaní Štatistického úradu SR*

### **KONTAKT**

[roman.pavelka@statistics.sk](mailto:roman.pavelka@statistics.sk)

**Boris VAŇO**  
**INFOSTAT – Výskumné demografické centrum**

## **VSTUP SLOVENSKÝCH ŽIEN DO REPRODUKCIE**

### **THE SLOVAK WOMEN'S ENTRY INTO REPRODUCTION**

#### **ABSTRAKT**

Vstup do reprodukcie významnou mierou ovplyvňuje intenzitu plodnosti, má však aj rad ďalších dôsledkov, napr. na zdravotný stav rodičiek a detí, stabilitu a životnú úroveň rodín. Z demografického hľadiska sa za vstup do reprodukcie považuje narodenie prvého dieťaťa. Článok hodnotí vstup do reprodukcie na úrovni medzinárodnej, celoštátnej aj regionálnej prostredníctvom priemerného veku žien pri prvom pôrode.

#### **ABSTRACT**

The entry into reproduction significantly affects the intensity of fertility, but it also has a number of other impacts, e.g. on the health of mothers and children, the stability and standard of living of families. From a demographic perspective, the birth of the first child is considered to be the entry into reproduction. The article evaluates the entry into reproduction at the international, national and regional level through the mean age of women at first birth.

#### **KLÚČOVÉ SLOVÁ**

reprodukcia, priemerný vek žien pri prvom pôrode, odkladanie pôrodov

#### **KEY WORDS**

reproduction, mean age of women at first birth, postponement of births

#### **1. ÚVOD**

O vstupe do reprodukcie sa hovorí pri rôznych príležitostiach. Najčastejšie v súvislosti so začiatkom sexuálneho spoluzitia ale napríklad aj v prípade asistovanej reprodukcie. Z demografického hľadiska sa za vstup do reprodukcie považuje narodenie prvého dieťaťa.

O optimálnom veku na vstup do demografickej reprodukcie sa dlhodobo vedú diskusie. Existuje široké spektrum názorov a pohľadov vyplývajúcich často aj z profesijného zamerania diskutujúcich. Lekár, demograf, sociálny pracovník i pedagóg majú svoj pohľad a svoje argumenty [1, 3, 4].

Skorší aj neskorší vstup do reprodukcie majú svoje výhody i nevýhody, preto nie je možné jednoznačne stanoviť vek, ktorý by sa dal označiť ako optimálny. Optimálnosť vstupu do demografickej reprodukcie významne závisí od konkrétnej životnej situácie príslušného páru. Vo všeobecnosti však platí, že by to nemalo byť príliš skoro ani príliš neskoro. Skorý vstup do reprodukcie je jeden z dôležitých predpokladov na vyššiu plodnosť (aj keď nie nevyhnutný). Na druhej strane mladí rodičia bývajú na rodičovstvo častejšie mentálne menej pripravení a mladé rodiny nezriedka materiálne nezabezpečené (čo sa týka príjmovej situácie alebo bývania). Tieto skutočnosti môžu mať negatívny vplyv na stabilitu rodín. Neskorý vstup do reprodukcie zas znižuje počet

narodených detí, prináša zdravotné riziká pre matku aj dieťa a zvyšuje riziko bezdetnosti.

Cieľom článku je poukázať na špecifiká vstupu do reprodukcie na Slovensku. Hodnotiť budeme vstup do reprodukcie na celorepublikovej úrovni (súčasťou bude aj hodnotenie postavenia SR v rámci Európskej únie) a všímať si budeme i regionálne rozdiely pri vstupe do reprodukcie na úrovni okresov SR. Ako ukazovatele využijeme priemerný vek žien pri prvom pôrode a to na celoštátnej úrovni, okresnej úrovni aj na úrovni EÚ. Ako doplnkový ukazovateľ na celoštátnej úrovni využijeme podiel detí narodených v manželstve do jedného roka od uzavretia manželstva.

## 2. METODICKÉ POZNÁMKY

Pri hodnotení vstupu do reprodukcie používame ako hlavný ukazovateľ priemerný vek žien pri prvom pôrode. Vo verejne dostupnej databáze Štatistického úradu SR DATAcube sú k dispozícii údaje o priemernom veku žien pri narodení prvého živého dieťaťa. Vo verejne dostupnej databáze Eurostatu sú k dispozícii údaje o priemernom veku žien pri narodení prvého dieťaťa (bez ohľadu na vitalitu). Keďže mŕtvorodenosť je na Slovensku i v Európe nízka, tieto dva údaje sa líšia len minimálne<sup>1</sup>. Preto v texte nebudeme rozlišovať, či ide o narodenie živého dieťaťa alebo všetkých detí a budeme hovoriť len o priemernom veku žien pri prvom pôrode.

Pri porovnaní údajov Štatistického úradu SR a Eurostatu o priemernom veku pri jednotlivých demografických udalostiach evidujeme určité rozdiely, ktoré vyplývajú zo spôsobu výpočtu. Týka sa to aj priemerného veku žien pri prvom pôrode. Štatistický úrad SR vychádza pri výpočte tohto ukazovateľa z počtu živonarodených detí podľa veku matky, Eurostat využíva na výpočet miery plodnosti podľa veku. Rozdiel nie je zásadný ale je viditeľný, pohybuje sa rádovo v desatinách roka.

Zmena evidencie narodených v zahraničí ženám s trvalým pobytom na Slovensku, ktorá sa uskutočnila na Slovensku v roku 2011, ovplyvnila všetky ukazovatele intenzity aj časovania plodnosti. Prejavila sa i na vývoji ukazovateľov použitých na hodnotenie vstupu do reprodukcie v tomto článku. Zlom v trende v roku 2012, ktorý je viditeľný na grafoch č. 1 a č. 2, má teda metodologické a nie obsahové pozadie.

Súčasnú úroveň vstupu do reprodukcie hodnotíme na základe najnovších dostupných údajov za rok 2022. Na hodnotenie vývoja sme zvolili obdobie od vzniku SR po súčasnosť, teda obdobie 1993 – 2022. Medzinárodné porovnanie je v článku spracované za obdobie 2013 – 2022. Začiatok tohto obdobia sme zvolili tak, aby sme sa vyhli zmene v evidencii narodených na Slovensku, pri voľbe konca obdobia sme zohľadnili dostupnosť najnovších údajov za krajiny EÚ 27 v databáze Eurostatu.

## 3. VSTUP ŽIEN DO REPRODUKcie NA SLOVENSKU

V druhej polovici 20. storočia patrilo Slovensko ku krajinám s najvyššou plodnosťou v Európe a s najnižším priemerným vekom žien pri pôrode. Jedným z najvýznamnejších predpokladov vysokej plodnosti bol práve skorý vstup do reprodukcie. Príčiny veľmi skorého vstupu obyvateľov do reprodukcie na Slovensku

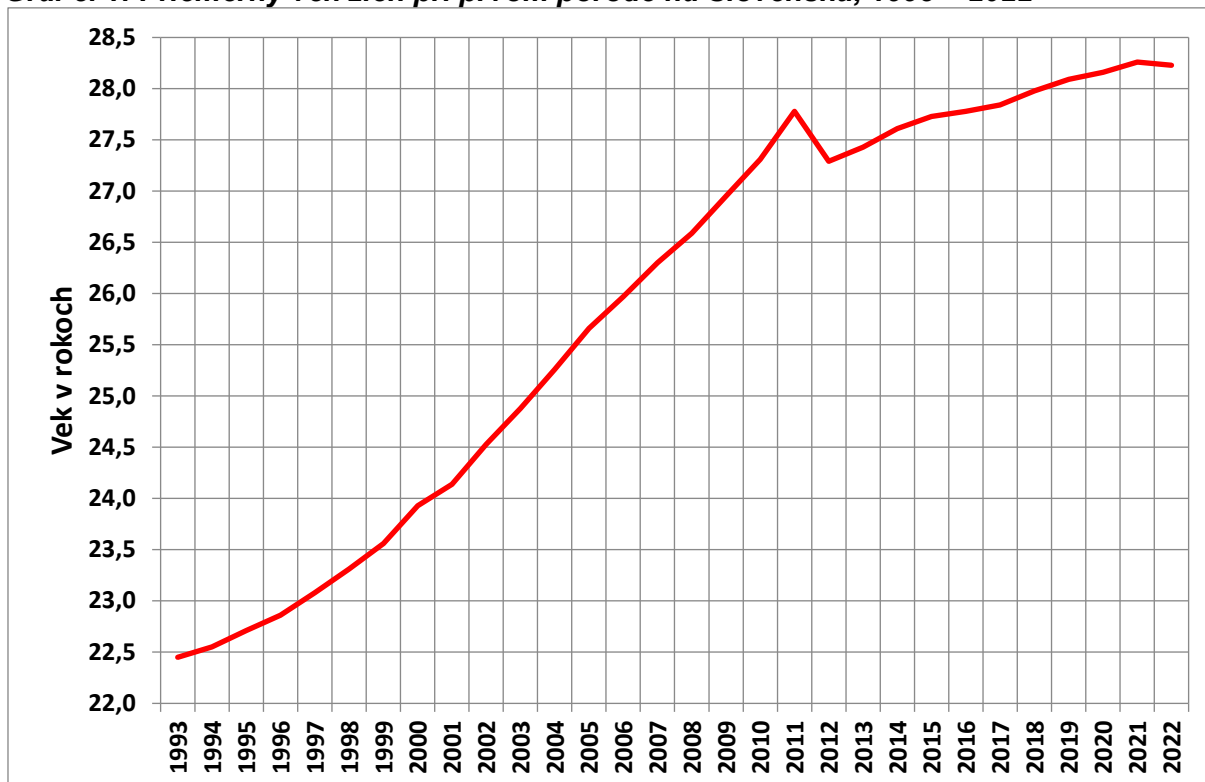
---

<sup>1</sup> Ako príklad minimálneho rozdielu medzi priemerným vekom žien pri pôrode všetkých detí a živonarodených detí uvádzame hodnoty za SR z roku 2022. Priemerný vek žien pri pôrode bol 29,3499 a priemerný vek žien pri pôrode živého dieťaťa bol 29,3503.

treba hľadať v tradíciách ale najmä v spoločensko-politickej situácii vo vtedajšom socialistickom Československu [6, 9].

Prvé náznaky odkladania pôrodov do vyššieho veku evidujeme na Slovensku už v 80. rokoch 20. storočia. Naplno sa tieto procesy prejavili po zmene politického režimu, s nástupom nového modelu reprodukčného správania v prvej polovici 90. rokov. Odkladanie pôrodov do vyššieho veku bolo jedným z hlavných znakov tohto modelu. Bolo tiež jedným z hlavných dôvodov výrazného zníženia plodnosti po roku 1990 [5, 6, 7, 8].

**Graf č. 1: Priemerný vek žien pri prvom pôrode na Slovensku, 1993 – 2022**



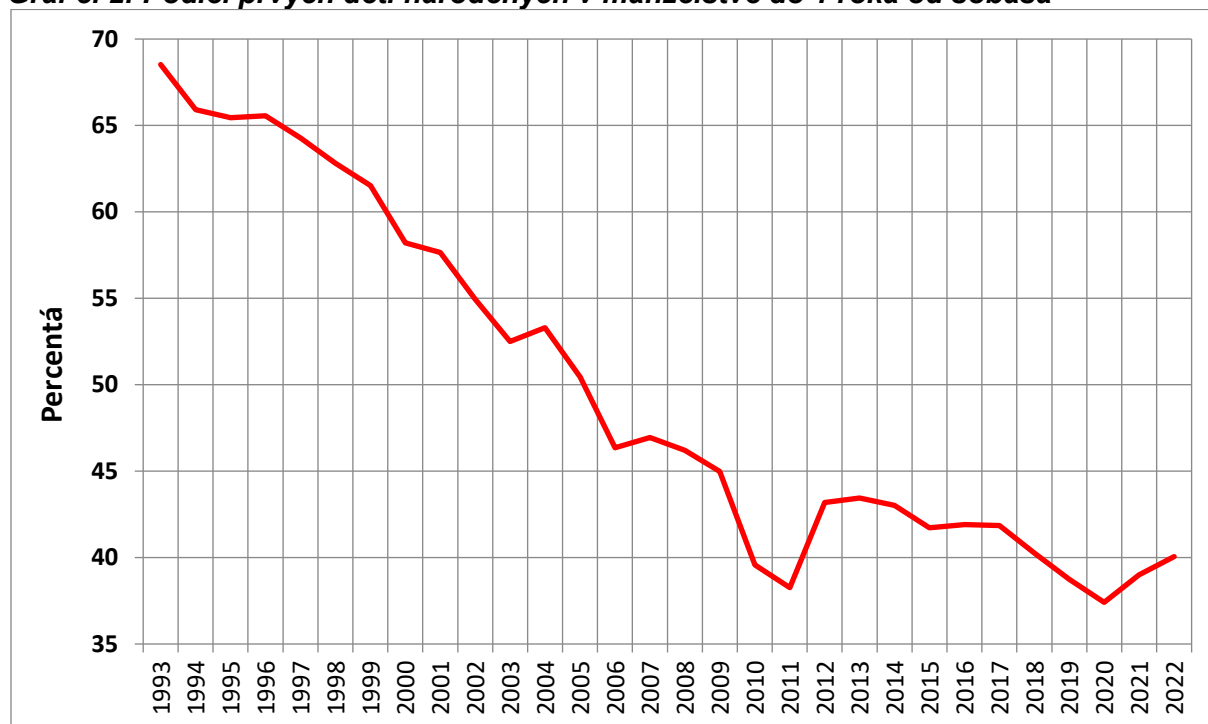
*Poznámka: Zlom v trende v roku 2012 je spôsobený zmenou evidencie narodených v zahraničí ženám s trvalým pobytom na Slovensku.*

**Zdroj: Štatistický úrad SR**

Najlepšie možno dokumentovať odklad vstupu do reprodukcie pomocou vývoja priemerného veku žien pri prvom pôrode (graf č. 1). Tento ukazovateľ zaznamenal nepretržitý rast od roku 1993 až po súčasnosť. Rok 2022 bol prvý od vzniku SR, keď sa priemerný vek žien pri prvom pôrode medziročne nezvýšil. Počas obdobia 1993 – 2022 evidujeme zvýšenie priemerného veku žien pri prvom pôrode z 22,4 roka na 28,3 roka, t. j. o 5,9 roka, resp. takmer 26 %. Medzi rokmi 2022 a 2023 sa priemerný vek žien pri prvom pôrode nepatrne znížil (o 0,1 roka). Výrazne prudší nárast priemerného veku žien pri prvom pôrode evidujeme v prvej časti sledovaného obdobia. Viac ako 90 % rastu tohto ukazovateľa pripadá na obdobie rokov 1993 – 2011 (graf č. 1). Príčinou bol hlavne proces rekuperácie, t. j. realizácie odložených pôrodov, ktorý sa začal na Slovensku po roku 2000 a ktorý kompenzoval silu odkladania pôrodov do vyššieho veku a spôsobil tak spomalenie rastu priemerného veku žien pri pôrode aj prvom pôrode [2, 6].

K podobným záverom ako pri hodnotení priemerného veku žien pri narodení prvého dieťaťa dospejeme aj pri hodnotení podielu narodených detí podľa času, ktorý uplynul od sobáša (graf č. 2). V období vysokej plodnosti sa na Slovensku rodila v manželstve väčšina prvých detí do jedného roka od sobáša. V 70. rokoch minulého storočia to bolo viac ako 60 % zo všetkých prvorodených detí v manželstve, v druhej polovici 80. rokov dokonca takmer 70 % [9].

**Graf č. 2: Podiel prvých detí narodených v manželstve do 1 roka od sobáša**



*Poznámka: Zlom v trende v roku 2012 je spôsobený zmenou evidencie narodených v zahraničí ženám s trvalým pobytom na Slovensku.*

**Zdroj: Štatistický úrad SR**

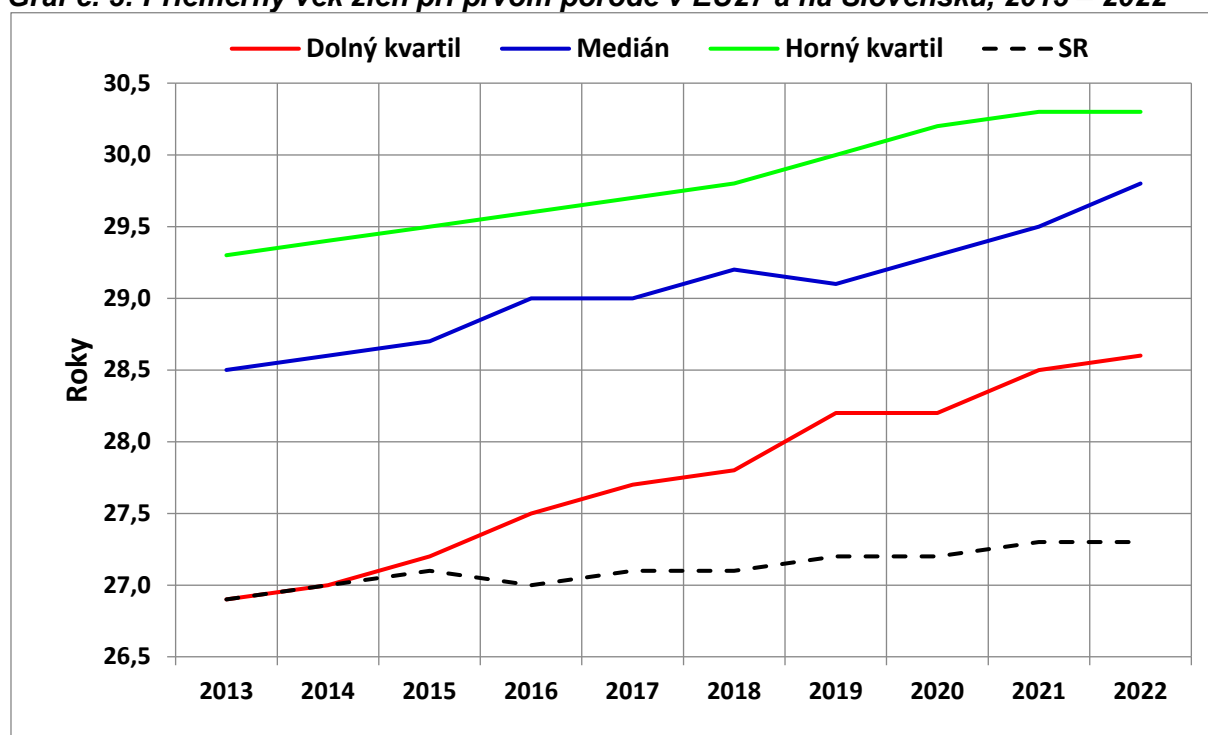
S nástupom nového reprodukčného modelu sa tesné prepojenie medzi sobášnosťou a plodnosťou znížilo. Ako sme už spomenuli, nový reprodukčný model je spojený s odkladom pôrodov do vyššieho veku, čo sa prejavilo aj na kontinuálnom poklese podielu detí narodených tesne po sobáši. Do roku 1995 evidujeme pokles prvých detí narodených v manželstve do jedného roka od sobáša na 65 % zo všetkých prvorodených detí v manželstve a do roku 2000 sa tento ukazovateľ znížil pod hranicu 60 %. Postupný pokles podielu prvých detí narodených do jedného roka od sobáša pokračoval až do súčasnosti, pričom súčasné hodnoty (na úrovni 40 %) sú v povojnovej histórii na Slovensku najnižšie (graf č. 2).

#### 4. VSTUP SLOVENSKÝCH ŽIEN DO REPRODUKČIE V EURÓPSKOM KONTEXTE

Odkladanie pôrodov do vyššieho veku je celoeurópsky trend. Vo všetkých krajinách EÚ sa za posledné desaťročia priemerný vek žien pri prvom pôrode zvýšil. Slovensko počas celého hodnoteného obdobia (roky 2013 – 2022) patrilo ku krajinám s najnižším priemerným vekom žien pri prvom pôrode, pričom postupne sa v rebríčku európskych krajín posúvalo smerom nadol. Prakticky po celý čas bol priemerný vek žien pri prvom pôrode na Slovensku pod hodnotou dolného kvartilu, pričom sa postupne od nej odkláňal a posúvalo smerom k minimálnej hodnote (graf č. 3).

Podľa údajov Eurostatu bol na začiatku hodnoteného obdobia priemerný vek žien pri prvom pôrode na Slovensku siedmy najnižší zo všetkých krajín EÚ 27. Nižšie hodnoty priemerného veku pri prvom pôrode ako Slovensko mali len ženy v pobaltských štátoch, Poľsku, Rumunsku a Bulharsku. V porovnaní s Bulharskom, v ktorom bol priemerný vek žien pri prvom pôrode najnižší zo všetkých krajín EÚ 27, bola hodnota na Slovensku o 1,2 roka vyššia. V roku 2013 bol priemerný vek žien pri prvom pôrode na Slovensku o 1,9 roka nižší ako priemer za EÚ 27. Za krajinami s najvyšším priemerným vekom žien pri prvom pôrode (Taliansko, Španielsko a Luxembursko) zaostávalo Slovensko o viac ako 3 roky. Do roku 2022 kleslo Slovensko v poradí krajín EÚ, čo sa týka priemerného veku žien pri prvom pôrode, na tretie miesto od konca (nižšie hodnoty priemerného veku žien pri prvom pôrode boli na konci hodnoteného obdobia len v Rumunsku a Bulharsku). Hodnota priemerného veku žien pri prvom pôrode na Slovensku bola v roku 2021 o 2,4 roka nižšia, ako bol priemer za EÚ 27 a rozdiel medzi Slovenskom a európskymi krajinami s najvyšším priemerným vekom žien pri prvom pôrode sa do roku 2022 zvýšil na viac ako 4 roky.

**Graf č. 3: Priemerný vek žien pri prvom pôrode v EÚ27 a na Slovensku, 2013 – 2022**



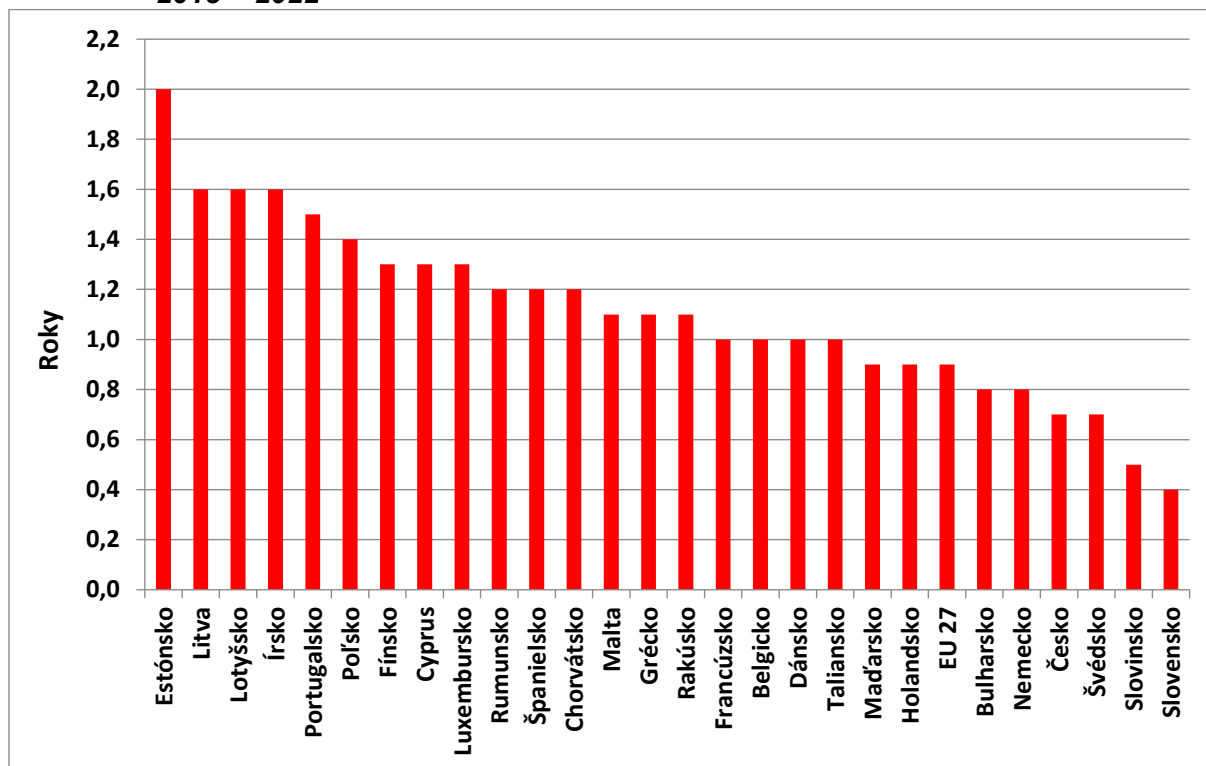
**Zdroj: Eurostat**

Priemerný vek žien pri prvom pôrode sa na Slovensku počas posledného desaťročia zvýšil najmenej zo všetkých krajín EÚ 27. Je to paradoxné, lebo krajiny s nízkou hodnotou tohto ukazovateľa (ku ktorým patrí aj Slovensko) majú väčší potenciál na jeho zvýšenie. V krajinách, ktoré mali v roku 2013 nízke hodnoty priemerného veku žien pri prvom pôrode evidujeme za posledné desaťročie najväčší rast tohto ukazovateľa (výnimku tvorí len Slovensko a Bulharsko). Okrem Slovenska a Bulharska evidujeme nízky nárast priemerného veku žien pri prvom pôrode ešte v Česku, Slovinsku, Nemecku a Švédsku (ide o krajiny na európske pomery s vysokou, resp. nadpriemernou hodnotou priemerného veku žien pri prvom pôrode). Vo všetkých týchto krajinách sa priemerný vek žien pri prvom pôrode zvýšil o menej ako 0,8 roka. Najvyšší rast priemerného veku žien pri prvom pôrode v období 2013 –



2022 (viac ako 1,4 roka) evidujeme v už spomínaných krajinách s nízkou hodnotou priemerného veku žien pri prvom pôrode (Estónsko, Litva, Lotyšsko, Poľsko) a ešte v Írsku a Portugalsku. S odstupom najväčšie zvýšenie priemerného veku žien pri prvom pôrode (o 2 roky) evidujeme za posledné desaťročie v Estónsku (graf č. 4).

**Graf č. 4: Zvýšenie priemerného veku žien pri prvom pôrode v krajinách EÚ 27, 2013 – 2022**



Zdroj: Eurostat

## 5. REGIONÁLNE ROZDIELY V ČASOVANÍ PRVÉHO PÔRODU

Regionálne rozdiely na Slovensku sú vo všeobecnosti výrazné a platí to aj v časovaní pôrodov. To, čo sme skonštatovali na nadnárodnej úrovni, vo všeobecnosti platí aj na regionálnej úrovni na Slovensku. To znamená, že v okresoch s vysokou plodnosťou evidujeme skorší vstup žien do reprodukcie a teda nižší priemerný vek žien pri prvom pôrode, a naopak. Zrejmý je aj trend odsúvania pôrodov do vyššieho veku vo všetkých okresoch, intenzita sa však v jednotlivých okresoch výrazne líši.

Medzi okresmi s nízkymi hodnotami priemerného veku žien pri prvom pôrode sú len okresy z východného Slovenska a severu a juhu stredného Slovenska. Naopak medzi okresmi s vysokými hodnotami sú len okresy zo západného Slovenska a z centrálnej časti stredného Slovenska (výnimkou sú košické mestské okresy – graf č. 5a, 5b, 5c).

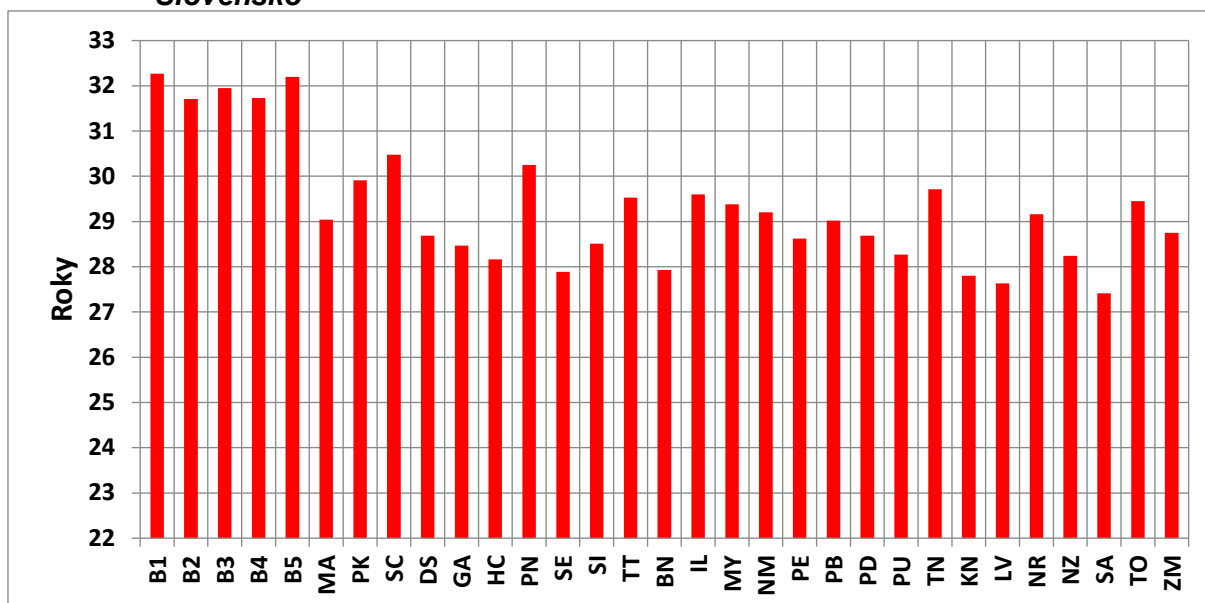
Priemerný vek žien pri prvom pôrode dosiahol v roku 2022 na celoštátnej úrovni hodnotu 28,2 roka. V jednotlivých okresoch sa priemerný vek žien pri prvom pôrode pohyboval od viac ako 32 rokov po menej ako 23 rokov. V 22 okresoch bol priemerný vek žien pri prvom pôrode veľmi blízko k priemernej hodnote za SR (ide o okresy vzdialené menej ako 0,5 roka od priemeru za SR). Väčšina z týchto „priemerných“ okresov sa nachádza na západnom a strednom Slovensku, z východoslovenských okresov patria do tejto skupiny len okresy Svidník a Humenné. Najbližšie hodnoty priemerného veku žien pri prvom pôrode k celorepublikovému priemeru evidujeme

v okresoch Kysucké Nové Mesto, Svidník, Púchov, Nové Zámky, Banská Štiavnica, Hlohovec a Turčianske Teplice (graf č. 5a, 5b, 5c).

Najvyšší priemerný vek žien pri prvom pôrode je v hlavnom meste. V jednotlivých bratislavských okresoch sa pohybuje od 31,7 roka v okrese Bratislava II až po okres Bratislava I, v ktorom ženy rodia svoje prvé dieťa v priemere vo veku 32,3 roka. Vyšší priemerný vek žien pri prvom pôrode ako 32 rokov je aj v okrese Bratislava V. Za bratislavskými okresmi nasledujú dva západoslovenské okresy (Senec a Piešťany) s priemerným vekom žien pri prvom pôrode nad 30 rokov. V 18 okresoch presahuje priemerný vek žien pri prvom pôrode hranicu 29 rokov, pričom v okresoch Košice I, Pezinok a Banská Bystrica sa približuje k hranici 30 rokov. Medzi okresmi s najvyšším priemerným vekom žien pri prvom pôrode (viac ako 29 rokov) sú len okresy zo západného a stredného Slovenska a tri košické mestské okresy (graf č. 5a, 5b, 5c).

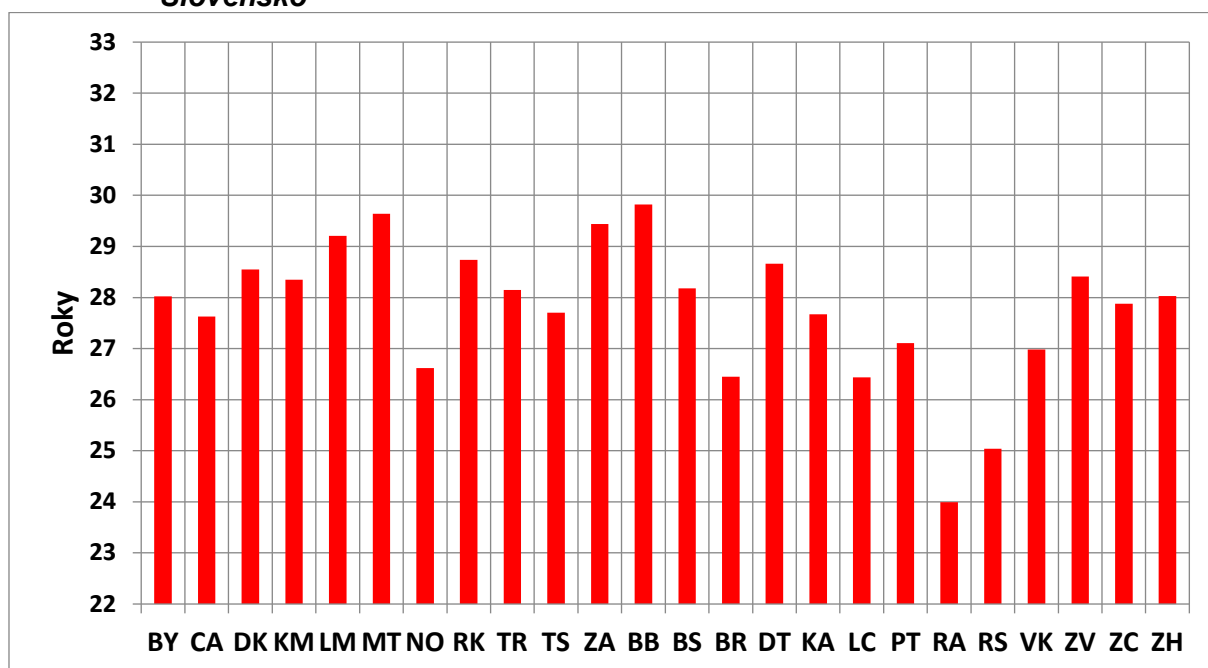
V 13 okresoch majú ženy pri prvom pôrode v priemere menej ako 26 rokov. Ide o okresy z východného Slovenska (pás okresov v západnej časti Prešovského a Košického kraja od poľských až po maďarské hranice) a stredoslovenský okres Revúca. Najnižší priemerný vek pri prvom pôrode majú ženy v okrese Gelnica (ide o jediný okres s nižšou hodnotou priemerného veku žien pri prvom pôrode ako 23 rokov). Hodnotu nižšiu ako 25 rokov dosahuje priemerný vek žien pri prvom pôrode ešte v okresoch Sabinov, Spišská Nová Ves, Kežmarok, Revúca a Medzilaborce, pričom v posledných dvoch menovaných okresoch bol priemerný vek žien pri prvom pôrode v roku 2022 nižší ako 24 rokov (graf č. 5a, 5b, 5c).

**Graf č. 5a: Priemerný vek žien pri prvom pôrode v okresoch SR v roku 2022 – západné Slovensko**



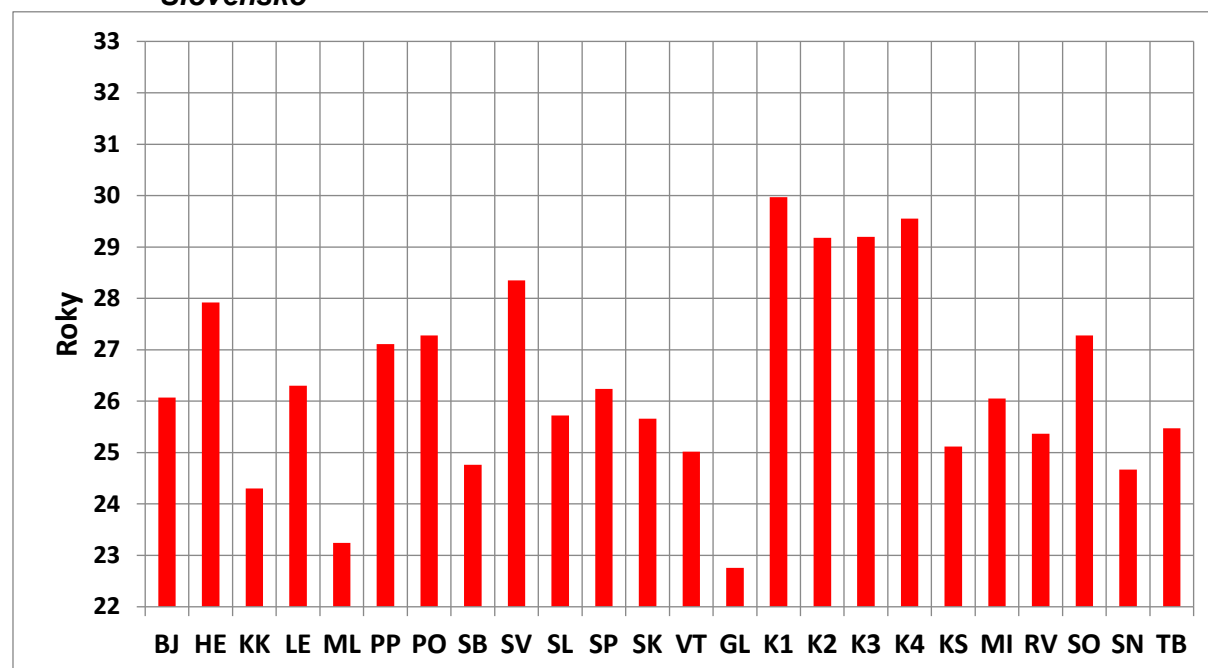
Zdroj: Štatistický úrad SR

**Graf č. 5b: Priemerný vek žien pri prvom pôrode v okresoch SR v roku 2022 – stredné Slovensko**



Zdroj: Štatistický úrad SR

**Graf č. 5c: Priemerný vek žien pri prvom pôrode v okresoch SR v roku 2022 – východné Slovensko**

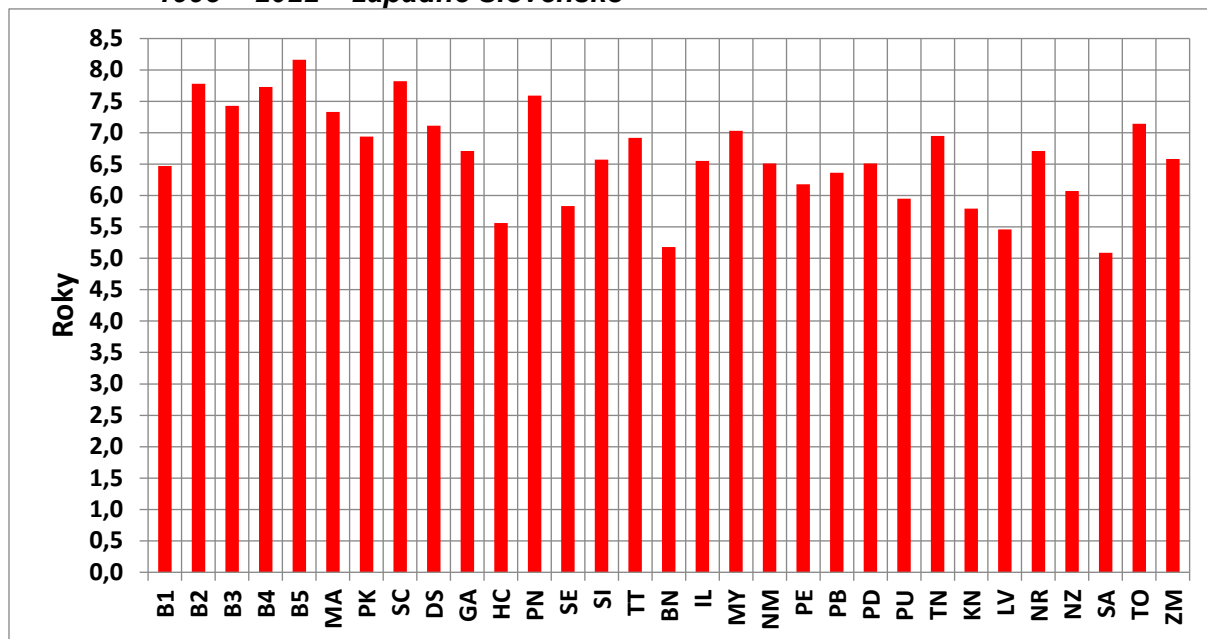


Zdroj: Štatistický úrad SR

Zvyšovanie priemerného veku žien pri prvom pôrode evidujeme za posledné tri desaťročia vo všetkých okresoch SR, intenzita zvyšovania však bola v jednotlivých okresoch rozdielna (graf č. 6a, 6b, 6c). Vo všeobecnosti možno konštatovať, že viac sa zvyšoval priemerný vek žien pri prvom pôrode v okresoch s vyššími hodnotami a menej v okresoch s nižšími hodnotami. To znamená, že výraznejšie zvýšenie priemerného veku žien pri prvom pôrode evidujeme na západe a v centrálnej časti stredného Slovenska a menšie na severe a juhu stredného Slovenska a na

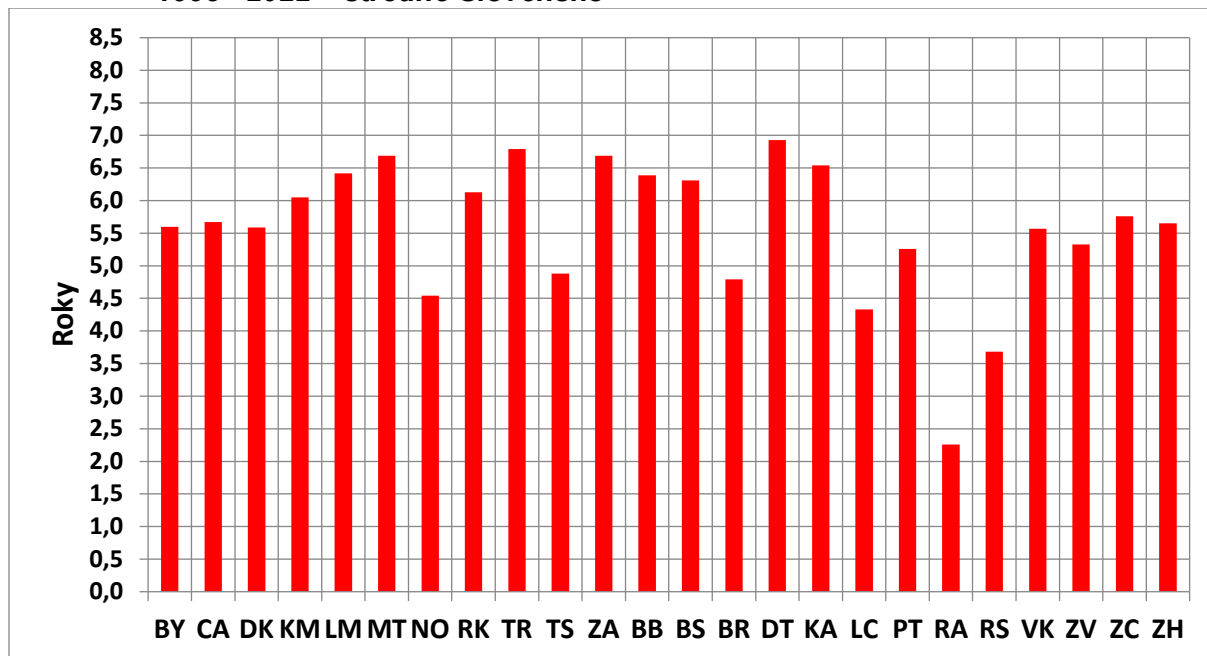
východnom Slovensku. Tým sa regionálne rozdiely, čo sa týka priemerného veku žien pri prvom pôrode, za posledných 30 rokov ešte viac prehĺbili.

**Graf č. 6a: Zmena priemerného veku žien pri prvom pôrode v okresoch SR za obdobie 1993 – 2022 – západné Slovensko**

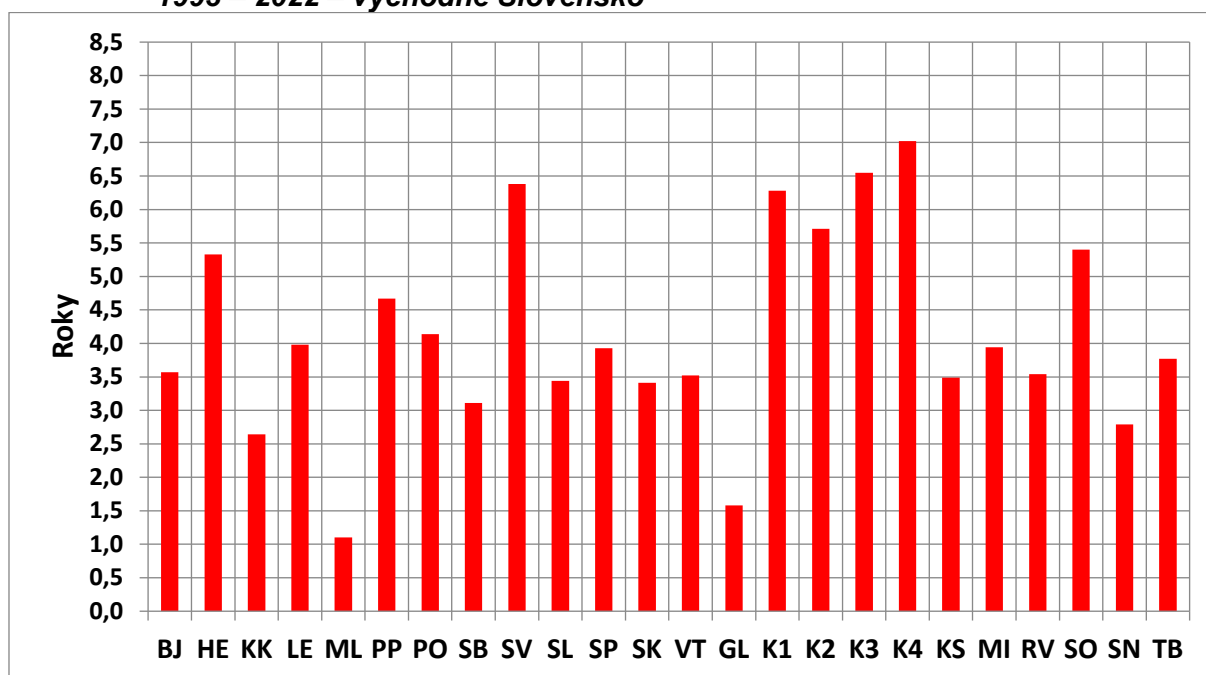


Zdroj: Štatistický úrad SR

**Graf č. 6b: Zmena priemerného veku žien pri prvom pôrode v okresoch SR za obdobie 1993 – 2022 – stredné Slovensko**



Zdroj: Štatistický úrad SR

**Graf č. 6c: Zmena priemerného veku žien pri prvom pôrode v okresoch SR za obdobie 1993 – 2022 – východné Slovensko****Zdroj: Štatistický úrad SR**

Zvýšenie priemerného veku žien pri prvom pôrode sa v období 1993 – 2022 pohybovalo v jednotlivých okresoch od 8,2 roka v okrese Bratislava V až po 1,1 roka v okrese Medzilaborce. Za okresom Bratislava V, ktorý je jediným okresom so zvýšením priemerného veku žien pri prvom pôrode o viac ako osem rokov, nasleduje desať okresov so zvýšením vyšším ako 7 rokov. Ide o zvyšné okresy Bratislavského kraja s výnimkou okresu Bratislava I, západoslovenské okresy Piešťany, Dunajská Streda, Topoľčany a Myjava a košický okres Košice IV (graf č. 6a, 6b, 6c).

Medzi okresmi s nadpriemerným zvýšením priemerného veku žien pri prvom pôrode sú prakticky len okresy zo západného a stredného Slovenska, výnimku tvoria iba košické mestské okresy a okres Svidník. Naopak, východoslovenské okresy výrazne prevládajú medzi okresmi s najmenším zvýšením priemerného veku žien pri prvom pôrode. Medzi dvadsiatkou okresov s najmenším nárastom tohto ukazovateľa sú z okresov mimo Prešovského a Košického kraja len dva okresy z juhu stredného Slovenska (Lučenec a Revúca) a jeden okres zo severu stredného Slovenska (Námestovo) (graf č. 6a, 6b, 6c).

Len minimálne zvýšenie priemerného veku žien pri prvom pôrode evidujeme v okrese Medzilaborce, nasleduje okres Gelnica, v ktorom sa priemerný vek žien pri prvom pôrode zvýšil za posledných 30 rokov približne o 1,5 roka. Zvýšenie o menej ako 3 roky nastalo ešte v okresoch Revúca, Kežmarok a Spišská Nová Ves. V ďalších 12 okresoch evidujeme zvýšenie priemerného veku žien pri prvom pôrode o menej ako 4 roky. Do desiatky okresov s najmenším zvýšením priemerného veku žien pri prvom pôrode patria, okrem už spomínaných piatich okresov s priemerným vekom žien pri prvom pôrode pod 3 roky ešte okresy Sabinov, Svidník, Stará Ľubovňa, Košice okolie a Vranov nad Topľou (graf č. 6a, 6b, 6c).

## 6. ZÁVER

Pre ženy na Slovensku je charakteristický pomerne skorý vstup do reprodukcie, a to napriek tomu, že už dlhšie obdobie sa posúva do vyššieho veku. Za posledné tri desaťročia sa priemerný vek žien pri prvom pôrode výrazne zvýšil, hoci rastúci trend sa po roku 2010 spomaľuje. Slovensko patrí dlhodobo ku krajinám s najnižším priemerným vekom žien pri prvom pôrode v EÚ. Zároveň sa na Slovensku zväčšujú regionálne rozdiely, čo sa týka vstupu žien do reprodukcie.

Napriek spomaleniu rastu priemerného veku žien pri prvom pôrode v poslednom desaťročí a náznakom jeho zastavenia v súčasnosti, je pravdepodobné, že vstup slovenských žien do reprodukcie sa bude naďalej posúvať do vyššieho veku, aj keď zvyšovanie priemerného veku žien pri prvom pôrode s určitou istotou nebude také intenzívne ako v 90. rokoch minulého storočia. Preto s veľkou pravdepodobnosťou budú ženy na Slovensku vstupovať do reprodukcie na európske pomery stále v pomerne nízkom veku.

## LITERATÚRA

- [1] BELLIELI, C.: The Best Age for Pregnancy and Undue Pressures. In: *Journal of family and reproductive health*, 2016, č. 3, s. 104 – 107.
- [2] BLEHA, B. – PILINSKÁ, V. – ŠPROCHA, B. – VAŇO, B.: Populačný vývoj Slovenska v prierezovom a kohortnom pohľade. Bratislava: INFOSTAT, 2022. 150 s. ISBN 978-80-89398-47-8.
- [3] HABBEMA, D.: Realizing of Desired Family Size: When Should Couples Start? In: *Human Reproduction*, 2015, č. 9, s. 2215 – 2221.
- [4] KOCOURKOVÁ, J. – SLABÁ, J. – ŠŤASTNÁ, A.: Neplánovaný odklad mateřství do vyššího věku v životní dráze českých žen. In: *Sociológia*, 2021, č. 1, s. 49 – 70.
- [5] KUČERA, T. – KUČEROVÁ, O. – OPARA, O. – SCHAICH, E. (eds.): *New Demographic Faces of Europe*. Berlin: Springer-Verlag, 2000.
- [6] POTANČOKOVÁ, M. – JURČOVÁ, D. – PILINSKÁ, V. – VAŇO, B.: Fertility between tradition and modernity. In: Frejka, T. – Sobotka, T. – Hoem, J. M. – Toulemon, L. (eds.): *Childbearing Trends and Policies in Europe. Demographic Research, Special Collection 7*, 2008, roč. 19, článok 25, s. 973 – 1018.
- [7] RABUŠIC, L.: Kde ty všechny děti jsou?. Praha: Sociologické nakladatelství, 2001.
- [8] VAN DE KAA, D. J.: Europe's Second Demographic Transition. In: *Population Bulletin*, 1987, č. 1, s. 1 – 57.
- [9] VAŇO, B.: Trendy populačného vývoja. In: Slovenská republika včera a dnes. Bratislava: Slovenská informačná agentúra, 1996. ISBN 80-8048-183-0, s. 89 – 96.

## RESUMÉ

Začiatok reprodukcie významnou mierou ovplyvňuje intenzitu plodnosti, má však aj rad ďalších dôsledkov, napr. na zdravotný stav rodičiek a detí, stabilitu a životnú úroveň rodín. Z demografického hľadiska sa za vstup do reprodukcie považuje narodenie prvého dieťaťa. Preto na hodnotenie vstupu do reprodukcie využívame ako hlavný ukazovateľ priemerný vek žien pri prvom pôrode.

Priemerný vek žien pri prvom pôrode sa na Slovensku dlhodobo zvyšuje. V období 1993 – 2022 evidujeme zvýšenie z 22,4 roka na 28,3 roka, t. j. o 5,9 roka, resp. 26 %. Extrémne prudký nárast priemerného veku žien pri prvom pôrode evidujeme od polovice 90. rokov 20. storočia až do roku 2010.

Odkladanie pôrodov do vyššieho veku je celoeurópsky trend. Vo všetkých krajinách EÚ sa za posledné desaťročia priemerný vek žien pri prvom pôrode zvýšil. Slovensko

dlhodobo patrí ku krajinám s najnižším priemerným vekom žien pri prvom pôrode, pričom postupne sa v rebríčku európskych krajín posúva smerom nadol.

Regionálne rozdiely na Slovensku sú výrazné a platí to aj o časovaní pôrodov. Vo všeobecnosti platí, že v okresoch s vysokou plodnosťou evidujeme skorší vstup žien do reprodukcie, a teda nižší priemerný vek žien pri prvom pôrode, a naopak. Zrejmy je aj trend odsúvania pôrodov do vyššieho veku vo všetkých okresoch, intenzita sa však v jednotlivých okresoch líši.

Medzi okresmi s nízkymi hodnotami priemerného veku žien pri prvom pôrode sú len okresy z východného Slovenska a severu a juhu stredného Slovenska. Naopak, medzi okresy s vysokými hodnotami patria len okresy zo západného Slovenska a z centrálnej časti stredného Slovenska (výnimkou sú košické mestské okresy).

Zvyšovanie priemerného veku žien pri prvom pôrode evidujeme za posledné tri desaťročia vo všetkých okresoch SR. Vo všeobecnosti možno konštatovať, že viac sa zvyšoval priemerný vek žien pri prvom pôrode v okresoch s vyššími hodnotami a menej v okresoch s nižšími hodnotami. To znamená, že regionálne rozdiely, čo sa týka priemerného veku žien pri prvom pôrode, sa za posledných 30 rokov ešte viac prehĺbili.

## RESUME

The entry into reproduction significantly affects the intensity of fertility, but it also has a number of other impacts, e.g. on the health of mothers and children, the stability and standard of living of families. From a demographic perspective, the birth of the first child is considered to be the entry into reproduction. That is why we use the mean age of women at the first birth as the main indicator for evaluating the entry into reproduction.

The mean age of women at first birth has been rising for a long time in Slovakia. In the period 1993-2022, we record an increase from 22.4 years to 28.3 years, i.e. by 5.9 years, or 26%. Extremely sharp increase in the mean age of women at first childbirth is registered from the mid-1990s to 2010.

Postponing childbirths to advanced age is a European trend. In all EU countries, the mean age of women at first birth has increased over the past decades. Slovakia has been one of the countries with the lowest mean age of women at first birth, while it is gradually moving downwards in the ranking of European countries.

Regional disparities in Slovakia are significant and this also applies to the timing of births. In general, in districts with high fertility, we record an earlier entry of women into reproduction and thus their lower mean age at first birth and vice versa. The trend of postponing births to older ages is also evident in all districts, but the intensity in individual districts varies.

Among the districts with low values of the mean age of women at first birth, there are only districts of eastern Slovakia and northern and southern part of central Slovakia. Conversely, the districts with high values include only districts from western Slovakia and from the central part of middle Slovakia (except the Košice city districts).

We have recorded an increase in the mean age of women at first birth over the last three decades in all districts of the Slovak Republic. In general, it can be stated that the mean age of women at first birth increased more in the districts with higher values and less in the districts with lower values. Thus, the regional disparities in the mean age of women at first birth have widened even more over the past 30 years.

## PROFESIJNÝ ŽIVOTOPIS

*Ing. Boris Vaňo* vyštudoval Vysokú školu ekonomickú v Bratislave, následne absolvoval postgraduálne štúdium z demografie na Univerzite Karlovej v Prahe. Od roku 1980 pracuje

*v Inštitúte informatiky a štatistiky ako výskumný pracovník v oblasti demografie. V rokoch 2000 – 2014 bol vedúcim Výskumného demografického centra. V období rokov 2006 – 2010 pôsobil ako podpredseda Slovenskej štatistickej a demografickej spoločnosti pre demografiu. Špecializuje sa na hodnotenie populačného vývoja, demografické prognózy a populačnú politiku.*

**KONTAKT**

[vano@infostat.sk](mailto:vano@infostat.sk)



Informatívny článok/Informative article

## VYUŽITIE VYMIEŇANÝCH MIKROÚDAJOV NA SPRESNENIE ŠTRUKTÚRY ÚDAJOV ŠTATISTIKY ZAHRANIČNÉHO OBCHODU S TOVAROM

### USE OF EXCHANGED MICRODATA TO REFINE THE DATA STRUCTURE OF FOREIGN TRADE IN GOODS STATISTICS

#### 1. ÚVOD

Od roku 2022 si členské štáty povinne vymieňajú mikroúdaje týkajúce sa ich intra-EÚ vývozov do ostatných členských štátov. Cieľom tejto výmeny údajov (ďalej len „MDE“ z MicroData Exchange), bolo vytvoriť podmienky na zníženie záťaže subjektov vykazujúcich v partnerskej krajine intra-EÚ dovozy. Čím väčší rozsah údajov prebratých z iných štátov by bol použitý pri zostavení slovenského intra-EÚ dovozu, tým viac subjektov, ktoré majú v súčasnosti povinnosť tieto dovozy vykazovať, by mohlo byť od tejto povinnosti oslobodených. Zároveň sa MDE údaje môžu použiť na spresnenie odhadov, ktoré sa v štatistike zahraničného obchodu (ŠZO) robia pre subjekty, ktoré nemajú povinnosť ich poskytovať.

Každý členský štát vrátane Slovenska, má povinnosť odoslať zozbierané údaje o intra-EÚ vývoze ostatným 26 členským štátom a Severnému Írsku do 30 dní po skončení referenčného mesiaca. V prípade, že poskytujúci štát získa dodatočné alebo opravené údaje, zašle ich formou aktualizácií. Údaje sa odosielajú do systému MDE-HUB v Eurostate, ktorý ich roztriedi podľa vykazanej partnerskej krajiny. Následne si každá z nich môže prevziať údaje, ktoré jej prináležia.

Aby bolo možné prevzaté MDE údaje použiť pri zostavovaní alebo spresňovaní údajov na strane dovozu, bolo potrebné na strane vývozu rozšíriť ich štruktúru o dve nové premenné. Prvou z nich je krajina pôvodu, podľa ktorej sa aj v slovenskej štatistike dovoz tovaru zverejňuje, druhou je identifikačné číslo pre daň z pridanej hodnoty partnerského subjektu (IČDPH partnera) v členskom štáte určenia, aby bolo možné v dovážajúcom štáte priradiť získané údaje konkrétnemu dovozcovi.

#### 2. ANALÝZA PREVZATÝCH ÚDAJOV

V rámci projektu financovaného Európskou úniou odbor štatistiky zahraničného obchodu analyzoval prevzaté MDE údaje. Prevzaté údaje boli porovnávané s údajmi, ktoré boli zozbierané alebo odhadnuté na Slovensku jednotlivo za všetkých 12 mesiacov roka 2022. Pozornosť bola zameraná najmä na vzájomné prepojenie IČDPH partnera s cieľom priblížiť štruktúru odhadovaných údajov na intra-EÚ dovoze zozbieraným údajom v ostatných členských štátoch, a to tak podľa partnerskej krajiny, ako aj podľa komodity na úrovni osemmiestnych podpoložiek kombinovanej nomenklatúry.

Prevzaté údaje boli analyzované nielen v termíne ich poskytnutia, ale aj po ich dodatočných aktualizáciách a porovnávané najmä s predbežnými slovenskými údajmi za celý rok 2022, ktoré boli zverejnené v marci 2023. Neskôr boli oba zdroje údajov analyzované znovu, pri príprave definitívnych údajov za rok 2022. Hodnota MDE údajov získaných štatistickými zisťovaniami v poskytujúcich krajinách sa aj po vylúčení údajov týkajúcich sa špecifických tovarov (elektrickej energie a zemného plynu v plynnom stave) pre jednotlivé referenčné mesiace od ich prvého poskytnutia

po posledné dostupné (v marci 2023) menila v rozpätí od -35,6 % (za referenčné obdobie apríl) do 47,1 % (za referenčné obdobie február). Najmenšia bola revízia júnových údajov, a to iba o 0,5 %.

Na podrobnejšie analýzy boli použité aktualizované MDE údaje, ktoré boli k dispozícii začiatkom marca 2023. V budúcnosti, keď budú vymieňané údaje v termíne ich poskytnutia v lepšej kvalite, bude ich možné využiť už skôr. Využiť sa však budú dať pravdepodobne až pri zostavovaní podrobných údajov o intra-EÚ dovoze tovaru, keďže v čase zasielania MDE údajov je už proces spracovania prvých (tzv. agregovaných) údajov ŠZO vo finálnej fáze.

Na základe analýz a porovnaní boli navrhnuté dve metódy na spresnenie štruktúry odhadovaných údajov ŠZO – konkrétne pre údaje týkajúce sa dovozu tovaru na diaľku (tzv. NOSS údaje) a pre údaje dopočítané za firmy, ktoré nemali povinnosť vykazovať intra-EÚ dovozy vzhľadom na ich nízku, tzv. podprahovú hodnotu.

### **3. VYUŽITIE PREVZATÝCH ÚDAJOV NA SPRESNENIE ŠTRUKTÚRY DOVOZOV TOVARU NA DIAĽKU (NOSS)**

Údaje o dovoze tovaru súkromnými osobami, ktoré si tovar nakúpili v krajinách EÚ prostredníctvom internetu, sú Štatistickému úradu SR poskytované Finančným riaditeľstvom SR v členení podľa krajiny zaslania a podľa štvrťrokov. MDE údaje sú v podrobnejšej štruktúre, poskytované na mesačnej báze a obsahujú aj informáciu o vykazujúcej krajine. Zároveň sú v nich pre dovozcov, ktorým nebolo pridelené identifikačné číslo pre DPH, alebo nie je známe, použité špeciálne zástupné identifikátory. Pre transakcie uvedeného typu by mal byť vykázaný druh obchodu 12 – Priamy obchod so súkromnými spotrebiteľmi/vykonávaný súkromnými spotrebiteľmi (vrátane predaja na diaľku).

Údaje o dovoze tovaru na diaľku sa osobitne pre každú partnerskú krajinu porovnávali s MDE údajmi, v ktorých bol vykázaný druh obchodu 12 a/alebo použitý špeciálny identifikátor pre partnera. Hľadala sa čo najlepšia zhoda hodnoty s prihladením na to, že v údajoch mohol byť chybné vykázaný druh obchodu alebo zástupný identifikátor. Podmnožina údajov, ktorých hodnota sa najviac približovala hodnote v NOSS údajoch, bola použitá ako podklad pre štruktúru na rozčlenenie údajov podľa komodity a štvrťročných údajov do jednotlivých mesiacov. Pre tri štáty s hodnotou spolu cca 9. mil. €, pre ktoré sa adekvátne údaje v MDE nevyskytovali, boli ako podklad použité tovary s najvyššou ročnou sumárnou hodnotou za všetky krajiny. NOSS údaje v hodnote cca 550 mil. € boli prvýkrát zahrnuté do definitívnych údajov ŠZO za rok 2022, a to v rovnakej granularite ako ostatné imputácie a dopočty.

### **4. VYUŽITIE PREVZATÝCH ÚDAJOV NA SPRESNENIE ŠTRUKTÚRY DOVOZOV (PODPRAHOVÉ ODHADY/DOPOČTY)**

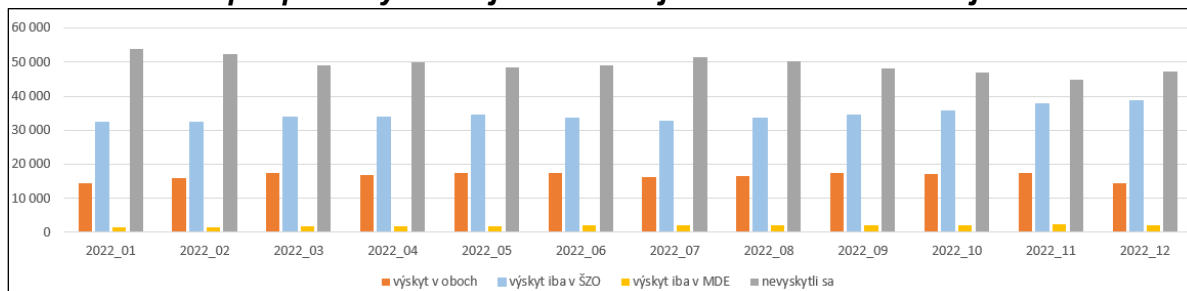
Pre subjekty, ktoré nemajú povinnosť vykazovať intra-EÚ dovozy, sa hodnoty dovozu odhadujú. V prípade, že sú k dispozícii hodnoty z daňových priznaní k DPH uvedené za daný referenčný mesiac ako nadobudnutie tovaru, použijú sa uvedené hodnoty. Ak pre dovážajúci subjekt za referenčný mesiac nie je hodnota k dispozícii, vypočíta sa štatistickými metódami na základe historických údajov. Hodnota za každý mesiac a subjekt sa rozdeľuje (tzv. disagreguje) do štruktúry podľa komodity a podľa krajiny zaslania. Pomocou koeficientov sa následne odvodí hmotnosti, prípadne množstvo v dodatkových merných jednotkách, ak existuje. Ako „vzor“ na disagregáciu

sa prioritne hľadá iný subjekt s rovnakou alebo aspoň podobnou ekonomickou činnosťou a/alebo čo najpodobnejšou hodnotou intra-EÚ dovozov.

Slovensko preberá MDE údaje z 26 členských štátov EÚ a Severného Írska, takže ak získa údaje napr. z Poľska, pre všetky zahrnuté transakcie je krajinou zaslania Poľsko. Keďže v prevzatých MDE údajoch sa vyskytujú ako slovenskí dovozcovia (IČ DPH partnera) aj také subjekty, ktoré na Slovensku nemajú povinnosť poskytovať údaje, je možné ich použiť priamo nielen v kombinácii s krajinou, ale aj s vykázanými komoditami, vrátane hmotnosti a množstva. Cieľom prepojenia dopočítaných podprahových údajov s MDE údajmi prostredníctvom zhodného identifikačného čísla (SK IČDPH) bolo zistiť, v akej miere sa hodnoty z oboch zdrojov zhodujú a v čo najväčšej miere využiť zozbierané MDE údaje na presnejšie rozdelenie údajov.

Údaje z oboch zdrojov boli porovnávané a analyzované za viac ako 100-tisíc subjektov, pre každý referenčný mesiac osobitne. Každý konkrétny subjekt sa v danom mesiaci v MDE buď vyskytol alebo nevyskytol. To platí rovnako aj v údajoch ŠZO, keďže subjekt mohol realizovať dovozy len v niektorých mesiacoch (graf. č. 1).

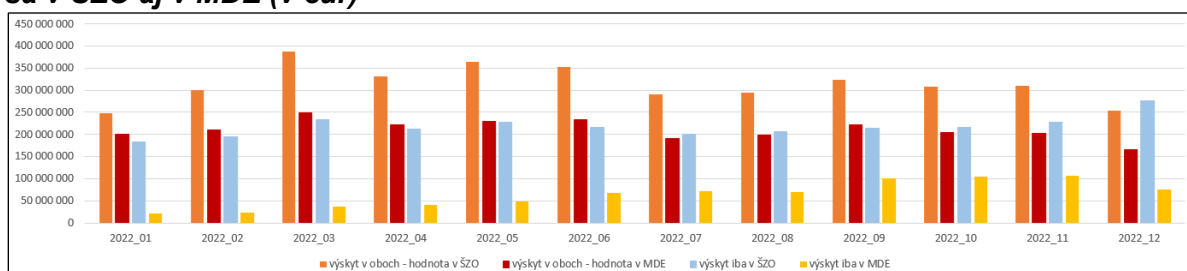
**Graf č. 1: Počet podprahových subjektov v údajoch ŠZO a v MDE údajoch**



**Zdroj údajov: Štatistický úrad SR, výpočty odboru štatistiky zahraničného obchodu k marcu 2023**

Priamo využiteľné boli údaje za subjekty, ktoré sa vyskytli v oboch súboroch v rovnakom mesiaci, a to v hodnote najmenej 1 euro (graf č. 2).

**Graf č. 2: Sumárna hodnota intra-EÚ dovozov podprahových subjektov vyskytujúcich sa v ŠZO aj v MDE (v eur)**



**Zdroj údajov: ŠÚ SR, výpočty odboru štatistiky zahraničného obchodu k marcu 2023**

Pre uvedené subjekty boli porovnávané hodnoty buď vzájomne zhodné, alebo jedna z hodnôt bola vyššia.

V celom rozsahu boli prebraté iba tie MDE údaje, ktorých hodnota bola menšia alebo sa rovnala údajom v ŠZO.

Uvedeným spôsobom bolo možné na základe MDE údajov pri príprave definitívnych údajov za rok 2022 presne špecifikovať údaje za viac ako 42-tisíc subjektov, ktorých hodnoty tvorili takmer štvrtinu (24,9 %) podprahových dopočtov na intra-EÚ dovoze.

Na spresnenie štruktúry údajov dovozcov, ktorých hodnota v ŠZO bola menšia ako hodnota zozbieraná na MDE, sa „preniesla“ krajinová a komoditná štruktúra MDE údajov pre viac ako 23-tisíc dovozcov. Uvedeným spôsobom bolo možné spresniť ďalších 8,8 % intra-EÚ dovozov „podprahových“ subjektov.

## 5. ZÁVER

Celkovo bolo v definitívnych údajoch štatistiky zahraničného obchodu s tovarom za rok 2022 využitím prevzatých MDE údajov spresnených 33,8 % údajov o intra-EÚ dovozoch tých subjektov, ktoré nemali povinnosť ich v uvedenom referenčnom období vykazovať. Zároveň sa spresnila aj štruktúra údajov týkajúcich sa dovozu tovaru na diaľku súkromnými osobami v roku 2022. Obdobným spôsobom sa aj naďalej so štvrťročnou periodicitou spresňujú údaje za rok 2023. Údaje tzv. podprahových intra-EÚ dovozcov budú spresnené pred publikovaním definitívnych údajov ŠZO za tento rok.

Odbor štatistiky zahraničného obchodu Štatistického úradu SR bude aj naďalej vymieňané MDE údaje analyzovať s cieľom rozšíriť ich využitie pri zostavovaní intra-EÚ dovozov nielen na zvýšenie kvality údajov, ale aj možného zníženia záťaže respondentov.

**Mgr. Emília MLADÁ**

*Autorka pracuje v odbore štatistiky zahraničného obchodu Štatistického úradu Slovenskej republiky.*

## Informácia/Information

### INTERAKTÍVNA PUBLIKÁCIA O BÝVANÍ V EURÓPE

#### INTERACTIVE PUBLICATION ABOUT HOUSING IN EUROPE

Na konci roka 2023 sa objavila na stránke Eurostatu aktualizácia zaujímavej interaktívnej publikácie **Housing in Europe – 2023 interactive publication** (<https://ec.europa.eu/eurostat/en/web/products-eurostat-news/w/wdn-20231130-1>), ktorá obsahuje zaujímavé údaje o rôznych aspektoch bývania v Európe.

Údaje za jednotlivé oblasti sú vo vizualizáciách prepojené s online databázami a komentáre k nim sa viažu na koniec novembra 2023, keď bola aktualizovaná interaktívna publikácia o bývaní v Európe spracovaná. Značná časť použitých databáz pochádza zo zisťovaní EU-SILC. Použité boli aj ďalšie databázy, týkajúce sa sledovania indexu cien nehnuteľností, cenovej hladiny podľa parity kúpnej sily, cien stavebných výrobcov, hrubej pridanej hodnoty podľa odvetví, tvorby hrubého fixného kapitálu, stavebných povolení a využívania pôdy podľa regiónov NUTS 2.

Za jednotlivé oblasti bývania sú k dispozícii stĺpcové a čiarové grafy podľa krajín EÚ a v prípade kartografov aj podľa regiónov. Stĺpcové grafy umožňujú bezprostredné porovnanie jednotlivých krajín s priemernou hodnotou daného ukazovateľa za celú EÚ v roku 2022. V ponúkaných čiarových grafoch je štandardne zobrazený vývoj hodnôt príslušného ukazovateľa za celú EÚ od roku 2010 do roku 2022, ale po zakliknutí konkrétnej krajiny sa objaví aj porovnanie vývoja hodnôt daného ukazovateľa v tejto krajine s priemerom za celú EÚ.

Publikácia má tri sekcie: **1. Ako žijeme**, **2. Náklady na bývanie** a **3. Výstavba**.

#### 1. AKO ŽIJEME

How we live 
House or flat – owning or renting
Size of housing
Quality of housing
Environmental impact of housing

V prvej časti publikácia poskytuje údaje o tom, či bývame v dome alebo byte, či nehnuteľnosť vlastníme alebo si ju prenajímame.

Veľkosť bývania sa meria ako priemerný počet izieb na osobu.

Kvalita bývania sa posudzuje podľa preplnenosti nehnuteľnosti, možnosti jej vykurovania a pod.

Vplyv bývania na životné prostredie sa posudzuje podľa objemu emisií skleníkových plynov z domácností na ich vykurovanie a chladenie.

Je všeobecne známe, že vlastnícka štruktúra nehnuteľností je v jednotlivých európskych krajinách značne diferencovaná. V roku 2022 bol najvyšší podiel vlastníctva nehnuteľností zaznamenaný v Rumunsku (95 % populácie žilo

v domácnosti, ktorá vlastnila nehnuteľnosť), na Slovensku (93 %), v Chorvátsku (91 %) a v Maďarsku (90 %). Na druhej strane najväčší podiel obyvateľov býva v prenájme v Nemecku (cca 53 %), nasledovalo Rakúsko (49 %) a Dánsko (40 %). V priemere viac ako dve tretiny obyvateľov EÚ (69 %) žili v domácnostiach, ktoré nehnuteľnosť vlastnili, zatiaľ čo zvyšných 31 % žilo v prenájme.

Zaujímavý je pohľad na európske krajiny, pokiaľ ide o typ bývania a lokalitu. V EÚ v roku 2022 žilo v priemere 52 % obyvateľov v dome, 47,5 % žilo v byte a 0,5 % bývalo v iných ubytovacích zariadeniach, ako sú hausbóty a pod. Na Slovensku žije zhruba 58 % obyvateľov v dome a zvyšných 42 % v byte. Najväčší podiel obyvateľov bývajúcich v domoch je v Írsku (takmer 90 %) a v bytoch v Španielsku (zhruba 66 %).

V mestách žilo v priemere 72 % obyvateľov EÚ v byte a 28 % v dome. V mestách a predmestiach to bolo 56 % ľudí žijúcich v dome a 43 % v byte, kým na vidieku 83 % obyvateľov žilo v dome a len 17 % v byte. Na Slovensku v mestách býva 15 % obyvateľov v domoch a 85 % v bytoch. V mestách a predmestiach je to pomer 43 % obyvateľov bývajúcich v domoch a 57 % v bytoch, kým na vidieku býva až 89 % obyvateľov Slovenska v domoch a len 11 % v bytoch.

V roku 2022 pripadalo v EÚ priemerne 1,6 izby na osobu. Spomedzi členských štátov najväčší počet izieb na osobu pripadal na Malte (2,3 izby na osobu) a na opačnom konci rebríčka boli Poľsko, Rumunsko a Slovensko (všetky 1,1 izby na osobu). Vo väzbe na uvedené pripadá na Slovensku až 3,1 osoby na jednu izbu, kým napr. vo Fínsku len 1,9 osoby na jednu izbu. V roku 2022 pripadalo v EÚ v priemere 2,3 osoby na jednu izbu.

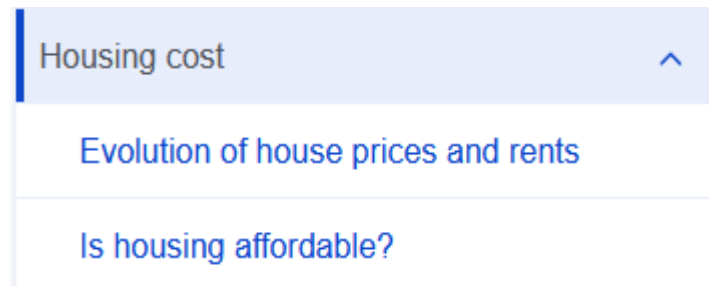
V dôsledku vyššieho počtu osôb pripadajúcich na jednu izbu vzniká preplnenosť nehnuteľností, čo vedie k znižovaniu kvality bývania. V EÚ v roku 2022 žilo v preplnených nehnuteľnostiach v priemere 16,8 % populácie, kým v roku ktorom? to bolo až 19,1 %. V roku 2022 bola najvyššia miera preplnenia nehnuteľností zaznamenaná v Lotyšsku (41,7 %) a najnižšia na Cypre (2,2 %). Na Slovensku sa miera preplnenosti nehnuteľností v roku 2022 znížila na 31,8 % z vyše 40 % v roku 2010.

Opakom preplnenej nehnuteľnosti je nedostatočne obývaná nehnuteľnosť, ktorá sa považuje za priveľkú pre potreby domácnosti, ktorá v nej žije. V EÚ v roku 2022 žila tretina obyvateľstva (33,6 %) v nedostatočne obsadenej nehnuteľnosti, pričom tento podiel sa od roku 2010 len nepatrne zvýšil. Aktuálne je najvyšší podiel nedostatočne obývaných nehnuteľností zaznamenaný na Malte (72,3 %) a najnižší v Rumunsku (7,3 %). Na Slovensku je aktuálne takmer 16 % nedostatočne obývaných nehnuteľností, kým v roku 2010 ich bolo vyše 11 %.

Kvalita bývania sa nehodnotí len podľa počtu osôb, ktoré bývajú v danej nehnuteľnosti, ale aj podľa toho, či v nej možno udržať primerané teplo, či je jej súčasťou vnútorné splachovacie WC, sprcha a vaňa. Sledujú sa tiež také vlastnosti nehnuteľnosti, ako spoľahlivosť strechy, či nie sú vlhké steny, podlahy alebo základy a pod. Detailné údaje o potenciálnych oblastiach nedostatkov v bývaní možno nájsť na <https://ec.europa.eu/eurostat/web/main/data/database> v sekcii Selected datasets, v časti Population and social condition, v podčasti Income and living condition, v skupine Material deprivation a v oddiele Housing deprivation.

Pokiaľ ide o environmentálny vplyv na bývanie, tak emisie skleníkových plynov z domácností na ich vykurovanie a chladenie sa pohybovali od 26 kg na obyvateľa vo Švédsku až po 1636 kg v Luxembursku. Priemer EÚ predstavuje 714 kg na obyvateľa a na Slovensku je to okolo 580 kg na obyvateľa.

## 2. NÁKLADY NA BÝVANIE



V tejto časti je prezentovaný vývoj cien nehnuteľností, cien nájomného a na porovnanie aj celková inflácia od roku 2010 do roku 2022. Hodnotí sa tiež dostupnosť bývania v mestách aj na vidieku.

V interaktívnej publikácii sú graficky sprístupnené cenové indexy, ktorých základom je priemer v roku 2015. Vývoj cien v oblasti bývania v roku 2022 v porovnaní s rokom 2010 získame podielom indexov za porovnávané roky.

Štandardne ponúknutý čiarový graf s hodnotami indexu cien nehnuteľnosti za celú EÚ môžeme interaktívne vymeniť za graf o vývoji cien nájomného alebo za graf o vývoji celkovej inflácie za celú EÚ. Následne môžeme pri všetkých uvedených možnostiach získať čiarové grafy porovnania vývoja hodnôt uvedených ukazovateľov za jednotlivé krajiny s priemerom EÚ.

Medzi rokmi 2010 a 2022 narástli ceny nehnuteľností v EÚ o 47 %. Vo väčšine členských štátov ceny nehnuteľností rástli. Najväčší nárast bol zaznamenaný v Estónsku (192 %), zatiaľ čo pokles zaznamenali v Taliansku (– 9 %) a na Cypre (– 5 %). Údaje za Grécko nie sú k dispozícii. Na Slovensku vzrástli ceny nehnuteľností za uvedené obdobie o 81,3 %.

Za obdobie 2010 až 2022 v EÚ narástlo nájomné o 18 %. K nárastu došlo vo všetkých členských štátoch okrem Grécka (– 25 %). Najväčší nárast zaznamenali v Estónsku (210 %), kým na Cypre narástlo nájomné len 0,2 %. Na Slovensku vzrástlo nájomné o 10,1 %, teda pozvoľnejšie, ako je priemer EÚ.

V rokoch 2010 až 2022 dosiahla inflácia v EÚ v priemere 28 %. Počas tohto obdobia bola ročná inflácia jednoznačne najvyššia v roku 2022 s 9,2 %. V uvedenom období rástla inflácia vo všetkých členských štátoch EÚ, najvýraznejšie v Estónsku (56 %). Najnižší nárast cien za uvedené obdobie zaznamenali v Grécku (12 %). Na Slovensku vzrástla inflácia v roku 2022 oproti roku 2010 o 36,3 %.

Za zmienku stojí porovnanie úrovne cien spojených s bývaním vrátane nákladov na vodu, elektrinu, plyn a iné palivá v jednotlivých krajinách k priemeru EÚ. V roku 2022 bola najvyššia úroveň cien spojených s bývaním v porovnaní s priemerom EÚ v Írsku (112 % nad priemerom EÚ) a najnižšia úroveň bola zaznamenaná v Poľsku (60 % pod priemerom EÚ). Slovensko patrí medzi päť krajín, v ktorých sa úroveň cien spojených s bývaním v porovnaní s priemerom EÚ v roku 2022 výrazne zvýšila oproti roku 2010 (z 56,1 % na 82,3 % pod priemerom EÚ).



Dostupnosť bývania je v prezentovanej interaktívnej publikácii posudzovaná pomocou viacerých ukazovateľov. Jedným z nich je miera preťaženia nákladov na bývanie (*Housing cost overburden rate*), ktorá predstavuje podiel obyvateľov žijúcich v domácnostiach, v ktorých celkové náklady na bývanie tvoria viac ako 40 percent ich disponibilného príjmu. V roku 2022 žilo v EÚ v takýchto domácnostiach 10,6 % obyvateľov miest, kým príslušná miera preťaženia nákladov na bývanie pre vidiecke oblasti bola 6,6 %. Najvyššia miera preťaženia nákladov na bývanie v mestách bola zaznamenaná v Grécku (27,3 %) a najnižšia na Slovensku (2,3 %). Vo vidieckych oblastiach bola najvyššia miera preťaženia nákladov na bývanie v Grécku (24,2 %) a najnižšia na Malte (0,2 %). Na slovenskom vidieku 2,8 %. Miera preťaženia nákladov na bývanie bola väčšinou vyššia v mestách a nižšia vo vidieckych oblastiach, ale v 7 členských štátoch to bolo naopak.

Ďalším spôsobom, ako zistiť, či je bývanie dostupné, je podiel nákladov na bývanie na celkovom disponibilnom príjme domácnosti. V roku 2022 v krajinách EÚ bolo v priemere 19,6 % disponibilného príjmu určených\*\*\* na náklady na bývanie. Najväčší podiel nákladov na bývanie z disponibilného príjmu je v Grécku (34,2 %) a najmenší na Malte (8,8 %). Na Slovensku predstavuje podiel nákladov na bývanie z disponibilného príjmu v priemernej domácnosti 15,2 %, v tzv. bohatšej domácnosti (s disponibilným príjmom vyšším ako 60 % národného mediánového príjmu) 13,4 % (v EÚ 16 %), ale v tzv. chudobnejšej domácnosti (s disponibilným príjmom nižším ako 60 % národného mediánového príjmu) až 26,1 % (v EÚ 37,9 %).

O (ne)dostupnosti bývania môžu dosť naznačiť i napríklad nedoplatky na hypotékach, nájomnom alebo účtoch za energie. Aj keď sa hlavne ceny nehnuteľností a nájomného v období od roku 2010 do roku 2022 zvýšili (ceny nehnuteľností o 47 %, ceny nájomného o 18 %), podiel ľudí žijúcich v domácnostiach s nedoplatkami na hypotékach, nájomnom alebo účtoch za energie v EÚ klesol z 12,4 % v roku 2010 na 9,2 % v roku 2022. Aktuálne najväčší podiel ľudí žijúcich v domácnostiach s uvedenými nedoplatkami zaznamenali v Grécku (45,5 %) a najmenší v Česku (2,7 %). Na Slovensku sa podiel uvedených domácností znížil z 12,1 % v roku 2010 na 8,3 % v roku 2022.

### 3. VÝSTAVBA

Construction	^
Construction sector	
Built-up areas	

Táto časť popri charakterizovaní sektoru stavebníctva a jeho vývoja od roku 2010 v krajinách EÚ zobrazuje aj najviac zastavané oblasti v Európe.

Podiel hrubej pridanej hodnoty (HPH) vytvorenej v stavebníctve na celkovej HPH sa v období 2010 až 2022 pohyboval v EÚ v priemere medzi 5 a 6 %. Najvyšší bol 5,8 % v roku 2010, potom mierne varioval a v roku 2022 dosiahol 5,5 %. Spomedzi členských štátov klesol podiel HPH v stavebníctve na celkovej HPH v rokoch 2010 až 2022 v 14 členských štátoch, pričom najväčší pokles bol v Španielsku, Bulharsku a na Cypre. V roku 2022 malo najväčší podiel HPH vytvorenej v stavebníctve na celkovej HPH Rakúsko (7,3 %) a najmenší Grécko (1,9 %). Na Slovensku je tento podiel aktuálne na úrovni 6,7 %, čo predstavuje mierny európsky nadpriemer.



Pre záujemcov o bývanie môže byť zaujímavou informáciou podiel investícií do bývania na celkovom HDP. V roku 2022 bol v tomto smere najštedrejší Cyprus (8,6 %) a najmenší podiel malo Grécko (1,6 %). Na Slovensku bol podiel investícií do bývania na celkovom HDP 4,3 %, rovnako ako v susednej Českej republike, čo je pod priemerom EÚ (5,9 %).

O veľkosti sektora stavebníctva výrazným spôsobom vypovedá počet podnikov, počet zamestnaných osôb a samotný rast zamestnanosti v tomto odvetví v jednotlivých krajinách i v regiónoch. Interaktívna mapa v prezentovanej publikácii obsahuje údaje za uvedené ukazovatele podľa jednotlivých európskych regiónov v členení NUTS 2. Zisťujeme, že napríklad spomedzi slovenských regiónov bolo v roku 2021 najviac stavebných podnikov aj zamestnancov v Západoslovenskom kraji (35 950 podnikov a 53 290 zamestnancov).

Významnou fázou procesu výstavby nehnuteľností na bývanie je získanie stavebného povolenia od orgánov verejnej moci. Všeobecne počet stavebných povolení v rámci EÚ hlavne od roku 2014 relatívne rástol vo väčšine členských štátov. Za obdobie rokov 2010 až 2022 to bolo najviac v Bulharsku (rast o 277 %) a naopak, najväčší pokles bol zaznamenaný v Taliansku (– 45 %). Na Slovensku sa v priebehu posledných desať rokov vydával takmer konštantný počet stavebných povolení (rast o 1,3 %).

Podľa údajov dostupných za rok 2018 sa v EÚ takmer 3 % pôdneho fondu využíva na obytné účely. V rámci regiónov je to však značne diferencované. Napríklad v regióne Berlína sa na tento účel využíva až 33 % pôdneho fondu. Na druhej strane, v 17 regiónoch bol tento podiel nižší ako 1,0 %, pričom najnižšie podiely zaznamenali v Aragónsku v Španielsku (0,2 %) a v Övre Norrland vo Švédsku (0,3 %). Na Slovensku sa na obytné účely používa pôdny fond v najväčšej miere v Bratislavskom kraji (3,2 %), nasleduje západné Slovensko (2,5 %), stredné Slovensko (2,4 %) a východné Slovensko (2,2 %).

Prezentovaná interaktívna publikácia poskytuje záujemcom na jednom mieste pomerne bohatý a ucelený komplex informácií v grafickej podobe o jednotlivých aspektoch bývania v jednotlivých krajinách EÚ. Ďalej poskytuje rôzne možnosti porovnávania jednotlivých ukazovateľov medzi krajinami aj vo vzťahu k priemeru celej EÚ. V prípade špecifickejšieho záujmu napr. o problematiku cien bývania je čitateľ usmernený zvlášť na metodiku zisťovania, na kompletnú databázu o štatistike cien bývania aj na dostupné publikácie z tejto oblasti.



Zdroj: <https://ec.europa.eu/eurostat/web/interactive-publications/housing-2023>

Aktuálna interaktívna publikácia je obsahovo v súlade s jej predchádzajúcimi edíciami, počínajúc rokom 2020. Posledná edícia umožňuje kontinuálne prezeranie grafov a komentárov od začiatku do konca celej publikácie, zatiaľ čo v predchádzajúcich edíciách bolo potrebné sa cez jednotlivé časti zvlášť preklikávať.

Interaktívne publikácie Eurostat vydáva od roku 2016. Takéto publikácie sú výborným nástrojom na propagovanie štatistiky i zvyšovanie štatistickej gramotnosti populácie.

**Ing. Mikuláš CÁR, PhD.**

*Autor vypracoval stále platnú metodiku zisťovania cien bývania v NBS, pracoval hlavne ako analytik sociálno-ekonomických javov a procesov a bol dlhoročným členom výboru Slovenskej štatistickej a demografickej spoločnosti.*

Nekrológ/Necrology

**ZA MARTINOU LUBYOVOU**

**IN MEMORY OF MARTINA LUBYOVÁ**  
**(\*12. 5. 1967 † 20. 11. 2023)**



*JUDr. Mgr. Martina Lubyová, PhD.*

V novembri 2023 niekoľko dní po výročí zamatovej revolúcie sme sa rozlúčili s **JUDr. Mgr. Martinou Lubyovou, PhD.** – významnou osobnosťou v oblasti slovenského školstva a vedy. V rôznych oficiálnych médiách sa o nej po jej smrti písalo ako o vedkyni, akademičke, ale tiež ministerke školstva, vedy, výskumu a športu Slovenskej republiky. Pre nás členov Katedry štatistiky Fakulty hospodárskej informatiky (FHI) Ekonomickej univerzity v Bratislave bola hlavne veselou, komunikatívnou a veľmi inteligentnou a ochotnou kolegyňou.

**Niektoré fakty z jej života**

Martina Lubyová sa narodila 12. mája 1967 v Bratislave v rodine známeho fyzika a bývalého dlhoročného predsedu Slovenskej akadémie vied Štefana Lubyho. Pokračovala v šľapajach otca a bola veľmi usilovná študentka o čom svedčí celý rad vedecko-pedagogických titulov, ktoré získala počas svojho života v rôznych študijných odboroch:

- 2-ročné mimoriadne štúdium žurnalistiky, Filozofická fakulta Univerzity Komenského v Bratislave, 1986 – 1988,
- Magisterka (Mgr.) v odbore biofyzika, Matematicko-fyzikálna fakulta Univerzity Komenského v Bratislave, 1986 – 1991,
- Magisterka (Mgr.) v odbore právo, Právnická fakulta Univerzity Komenského v Bratislave, 1991 – 1997,
- PhD. v odbore štatistika, Fakulta hospodárskej informatiky Ekonomickej univerzity v Bratislave, 1998 – 1999, dizertačná práca Empirické skúmania vývoja situácie na trhu práce v Slovenskej republike počas prechodového obdobia,
- Juris Utrisque doctor (JUDr), Právnická fakulta Univerzity Komenského v Bratislave, 1998 – 1999, rigorózna práca v oblasti práva sociálneho zabezpečenia EÚ a SR,
- PhD. v odbore ekonomika (Ph.D. in Economics), State University of New York (SUNY) a Centre for Economic Research and Graduate Education (CERGE) Univerzity Karlovej v rokoch 1991 – 2002, dizertačná práca v oblasti vývoja trhu práce tranzitívnych ekonomík SR a ČR.

Martina Lubyová počas svojho života prešla mnohými pracovnými pozíciami nielen na národnej, ale aj medzinárodnej úrovni. Jej pôsobenie ako vedkyne sa začalo v Prahe na Českej akadémii vied a to v Národohospodárskom ústave v rokoch 1993 až 1995. Následne od roku 1995 pracovala ako výskumná pracovníčka v Slovenskej akadémii vied v Bratislave a jej domovským pracoviskom bol Prognostický ústav.

Martina bola nielen vedkyňa, ale aj pedagogička. Na svoju pedagogickú kariéru nastúpila takisto v Prahe. Na Univerzite Karlovej pôsobila v odbore štatistika a ekonometria ako aj v odbore čistá teória medzinárodného obchodu súbežne so svojím pôsobením v Českej akadémii vied. Na Ekonomickej univerzite v Bratislave, na Fakulte hospodárskej informatiky pokračovala v pedagogickej činnosti jednak počas svojho doktorandského štúdia a následne vo vedľajšom pracovnom pomere. V odbore štatistiky sa zameriavala predovšetkým na teóriu výberových zisťovaní a sociálnu štatistiku.

Z medzinárodného hľadiska za najvýznamnejšiu možno považovať jej pozíciu riaditeľky Medzinárodnej organizácie práce (ILO), Subregionálneho úradu pre východnú Európu a Strednú Áziu so sídlom v Moskve, kde pôsobila od roku 2000 až do roku 2010. Bola špecialistkou na rozvoj zamestnanosti, pričom v roku 1995 pôsobila tiež v Paríži ako profesionálna pracovníčka v rámci programu Partners in Transition, ktorý bol zameraný na vstup krajín V4 do OECD.

Po svojom návrate zo zahraničia pracovala opäť ako vedkyňa i pedagogička v Slovenskej akadémii vied (napr. aj v CE CESTA – Centre excelentnosti: Centre strategických analýz SAV) ako aj na Fakulte hospodárskej informatiky Ekonomickej univerzity v Bratislave až do roku 2017. V septembri roku 2017 bola vymenovaná na pozíciu ministerky školstva. Ako ministerka dosiahla mnohé úspechy, napríklad zvýšenie plátov učiteľov o niekoľko desiatok percent, zároveň to však bolo obdobie plné náročných úloh a stresu.

### **Moja skúsenosť s Martinou**

Martina bola moja rovesníčka, takže ma správa o jej úmrtí veľmi zasiahla. Človek nie je nikdy dostatočne pripravený na smrť niekoho blízkeho a to platí aj v prípade, keď ide o blízku kolegyňu, s ktorou som ešte pred jej ministerskou kariérou úzko spolupracovala. Martina mala neuveriteľne veľký rozhľad a organizačné schopnosti, ktoré jej umožnili v čase pôsobenia na Katedre štatistiky Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave zapojiť našu fakultu do projektu medzinárodného významu. Išlo o 7. rámcový projekt Európskej únie pod názvom LLLIGHT in Europe: Lifelong Learning, Innovation, Growth and Human capital Tracks in Europe (Celoživotné vzdelávanie, inovácia, rast a toky ľudského kapitálu v Európe), vďaka ktorému katedra a jej členovia získali nové praktické skúsenosti, ale tiež uznanie od vedenia univerzity. Martinu som si veľmi vážila za jej prístup a súčasne ľahkosť pri riešení rôznych problémov. Ešte aj v poslednom roku svojej ťažkej choroby prišla na univerzitu, plná optimizmu s cieľom byť nápomocná pri akreditácii študijných programov na FHI.

**Som vďačná, že mi život doprial stretnúť sa s osobnosťou, ako si bola Ty.  
Čeť Tvojej pamiatke!**

*Za Katedru štatistiky Fakulty hospodárskej informatiky Ekonomickej univerzity v Bratislave*  
**doc. Ing. Mária Vojtková, PhD.**

**PRIPRAVUJEME/COMING SOON**

**Andrej CHROMEČEK**

BICYKEL A KOLOBEŽKA AKO PROSTRIEDKY DOCHÁDZKY DO ZAMESTNANIA  
PODĽA SODB 2021

BICYCLE AND SCOOTER AS A MEANS OF COMMUTING TO WORK ACCORDING  
TO THE 2021 POPULATION AND HOUSING CENSUS

\* \* \*

**ONLINE VERZIA ČÍSLA 2/2024 SLOVENSKEJ ŠTATISTIKY A DEMOGRAFIE JE  
VEREJNE DOSTUPNÁ** na internetovej stránke [slovak.statistics.sk](http://slovak.statistics.sk) a [ssad.statistics.sk](http://ssad.statistics.sk)  
od 15. APRÍLA 2024.

**THE ONLINE VERSION OF THE JOURNAL SLOVAK STATISTICS AND  
DEMOGRAPHY No 2 (2024) IS PUBLICLY BE AVAILABLE** at the website  
[slovak.statistics.sk](http://slovak.statistics.sk) and [ssad.statistics.sk](http://ssad.statistics.sk) from **APRIL 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 nájdete autori na [ssad.statistics.sk](http://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](http://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](http://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](http://ssad.statistics.sk).

## SLOVENSKÁ ŠTATISTIKA A DEMOGRAFIA

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

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

### **Vydáva:**

Štatistický úrad Slovenskej republiky  
Slovenská štatistická a demografická spoločnosť

### **Identifikačné čísla vydavateľov:**

IČO 00166197 / 00178764

### **Vychádza:**

Štyrikrát ročne

### **Dátum vydania:**

15. apríl 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

## SLOVAK STATISTICS AND DEMOGRAPHY

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

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

### **Issued by:**

Statistical Office of the Slovak Republic  
Slovak Statistical and Demographic Society

### **Companies registration numbers:**

00166197 / 00178764

### **Published:**

Four times a year

### **Date of issue:**

15<sup>th</sup> April 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