

Štatistický úrad Slovenskej republiky
The Statistical Office of the Slovak Republic

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

vedecký časopis/scientific journal

4/2023
ročník 33



ŠTATISTICKÝ
ÚRAD
SLOVENSKEJ
REPUBLIKY

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

SLOVENSKÁ ŠTATISTIKA A DEMOGRAFIA

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

Zahraniční poradcovia/Foreign Consultants

Gabriela Czanner

University of Liverpool
Veľká Británia/United Kingdom

Jitka Langhamrová

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

Estefanía Mourelle Espasandín

Universidade da Coruña
Španielsko/Spain

Michaela Potančoková

Joint Research Centre,
European Commission
Taliansko/Italy

Hana Řezanková

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

Milan Stehlík

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

Výkonná redaktorka/Executive Editor

Silvia Hudecová

Jazykové redaktorky/Language Editors

Slovenský jazyk/Slovak Language

Silvia Duchková

Anglický jazyk/English Language

Andrea Okenková

SLOVAK STATISTICS AND DEMOGRAPHY

The scientific peer-reviewed journal founded in 1991. Individual copies of the journal are available to readers in electronic form at the websites ssad.statistics.sk and slovak.statistics.sk. The opinions of the authors do not necessarily correlate with the opinions of the publisher.

Redakčná rada/Editorial Board

Ľudmila Ivančíková

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

Mikuláš Cár

Slovenská štatistická a demografická spoločnosť
Slovak Statistical and Demographic Society

Helena Glaser-Opitzová

Štatistický úrad SR
Statistical Office of the SR

Ján Haluška

INFOSTAT Bratislava

Iveta Stankovičová

Univerzita Komenského v Bratislave
Comenius University in Bratislava

Erik Šoltés

Ekonomická univerzita v Bratislave
University of Economics in Bratislava

Pavol Tišliar

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

Boris Vaňo

INFOSTAT - Výskumné demografické centrum
INFOSTAT - Demographic Research Centre

Adresa redakcie/Address of Editorial Office

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

E-mailová adresa/E-mail address

SSaD@statistics.sk

ssad.statistics.sk
www.statistics.sk

OBSAH/CONTENTS

I. VEDECKÉ ČLÁNKY/SCIENTIFIC ARTICLES

Boris VAŇO 2
OBYVATEĽSTVO SLOVENSKEJ REPUBLIKY PODĽA RODINNÉHO STAVU
POPULATION OF THE SLOVAK REPUBLIC BY MARITAL STATUS

Milan TEREK 20
CHARAKTERISTIKY VÝKONNOSTI PROCESU V METODOLÓGII SIX SIGMA
PROCESS PERFORMANCE CHARACTERISTICS IN SIX SIGMA METHODOLOGY

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

Roman PAVELKA 41
ÚVOD DO STATISTICKÉHO MODELOVÁNÍ POMOCÍ ANALYTICKÉHO SYSTÉMU
SAS
INTRODUCTION TO STATISTICAL MODELLING USING THE SAS ANALYSIS
SYSTEM
Informatívny článok/Informative article

Albert IVANČÍK, Helena GLASER-OPITZOVÁ 62
TRETIE KOLO PEER REVIEW 19. – 23. jún 2023
THIRD ROUND OF PEER REVIEW 19. – 23. June 2023
Informácia/Information

Martin BOĎA, Rudolf ZIMKA 65
AMSE 2023 – 25. ROČNÍK KONFERENCIE
AMSE 2023 – 25TH YEAR OF THE CONFERENCE
Informácia/Information

III. PRIPRAVUJEME/COMING SOON 68

Boris VAŇO
INFOSTAT – Výskumné demografické centrum

OBYVATEĽSTVO SLOVENSKEJ REPUBLIKY PODĽA RODINNÉHO STAVU

POPULATION OF THE SLOVAK REPUBLIC BY MARITAL STATUS

ABSTRAKT

Rodinný stav je jednou z charakteristík obyvateľstva, ktoré sa zisťujú výlučne v sčítaní obyvateľov. Analýza v článku je zameraná na hodnotenie vývoja rodinného stavu obyvateľov Slovenska za posledné tri desaťročia s využitím údajov z posledných štyroch sčítaní obyvateľov a na hodnotenie štruktúry a regionálnych rozdielov v rodinnom stave obyvateľov na základe výsledkov posledného sčítania obyvateľov z roku 2021.

ABSTRACT

Marital status is one of the characteristics of population which is obtained exclusively in the population census. The analysis in the article is aimed at evaluating the development of the marital status of the Slovak population over the past three decades using data from the last four population censuses and at evaluating the structure and regional differences in the population's marital status based on the results of the last population census from the year 2021.

KLÚČOVÉ SLOVÁ

rodinný stav, sčítanie obyvateľov, reprodukčné správanie

KEY WORDS

marital status, population census, reproductive behaviour

1. ÚVOD

Rodinný stav je významná charakteristika obyvateľstva. Rodinný stav obyvateľov vplýva na rôzne stránky spoločenského života a dôležitý je aj pre vývoj reprodukčného správania obyvateľov, pretože reprodukcia je do značnej miery viazaná na rodinu, resp. partnerské spolužitie.

Rodinný stav obyvateľov sa určuje vzhľadom na manželstvo a v konečnom dôsledku závisí od sobášnosti, rozvodovosti a úmrtnosti. V pozadí však stojí veľké množstvo faktorov, ktoré rodinný stav ovplyvňujú sprostredkované cez reprodukčné faktory. Najčastejšie ide o faktory legislatívne, kultúrne a ekonomické.

Slabým miestom tejto dôležitej charakteristiky je skutočnosť, že nepostihuje kohabitácie. Osoby, ktoré žijú v partnerskom zväzku, sa vzhľadom na rodinný stav považujú za slobodné. Pritom osoby žijúce v kohabitácii majú, čo sa týka reprodukčného správania, obvykle bližšie k osobám žijúcim v manželstve ako ku slobodným.

Cieľom článku je komplexné hodnotenie rodinného stavu obyvateľov na Slovensku na základe údajov zo sčítaní obyvateľov, domov a bytov (ďalej ako „SODB“). Analýza rodinného stavu obsahuje hodnotenie vývoja, štruktúry aj regionálnych rozdielov. Zmeny rodinného stavu v čase budeme hodnotiť počas obdobia 1991 – 2021.

Regionálne rozdiely budeme hodnotiť na úrovni krajov SR v roku 2021. Na hodnotenie štrukturálnych rozdielov v roku 2021 využijeme faktory, ktoré významne ovplyvňujú reprodukčné správanie obyvateľov a tým aj rodinný stav obyvateľov – pohlavie, vek, počet živonarodených detí, vzdelanie, národnosť a náboženské vyznanie [1, 2, 4].

2. ÚDAJE A METODICKÉ POZNÁMKY

Všetky údaje, použité na hodnotenie rodinného stavu obyvateľov na Slovensku, pochádzajú z SODB. Vývoj rodinného stavu obyvateľov hodnotíme na základe údajov z posledných štyroch SODB (z rokov 1991, 2001, 2011 a 2021). Štrukturálne a regionálne rozdiely v rodinnom stave analyzujeme na základe údajov z najnovšieho SODB z roku 2021 [6, 7, 8, 9].

Rodinný stav obyvateľov hodnotíme na základe rozdielov v čase (za obdobie 1991 – 2021), regionálnych rozdielov v roku 2021 a rozdielov v štruktúre podľa jednotlivých sledovaných ukazovateľov v roku 2021. Pokiaľ hodnotíme rodinný stav samostatne (či už na celoštátnej, alebo regionálnej úrovni), t. j. bez kombinácie s iným ukazovateľom, využijeme na hodnotenie údaje o celej populácii. Na hodnotenie rodinného stavu v kombinácii s iným ukazovateľom (v našom prípade pohlavím, vekom, vzdelaním, počtom detí, národnosťou a religiozitou) využijeme údaje o obyvateľoch vo veku od 15 rokov. Vynechaním detskej zložky obyvateľstva pri týchto analýzach sa snažíme eliminovať vplyv tej časti populácie, ktorej sa tieto štrukturálne faktory priamo netýkajú, t. j. osoby, ktoré nemôžu mať iný rodinný stav ako slobodný, resp. slobodná, nemôžu mať ukončené žiadne vzdelanie, okrem výnimiek nemajú deti a v sčítaní obyvateľov nemôžu osobne deklarovat' národnosť ani príslušnosť k cirkvi.

V tabuľkách sú uvedené len absolútne počty obyvateľov, relatívne údaje sa nachádzajú v grafoch. V tabuľkách sa nachádzajú aj osoby s nezistenými údajmi. Pri výpočte jednotlivých štruktúr sa však nezistené údaje nebrali do úvahy. To znamená, že do hodnotenia štruktúry rodinného stavu obyvateľov SR sú zahrnuté len tie osoby, o ktorých sa v sčítaní obyvateľov zistili všetky príslušné údaje. Keďže o každej sčítanej osobe je k dispozícii údaj o pohlaví, veku a mieste trvalého pobytu, ide o osoby s nezisteným rodinným stavom, počtom živonarodených detí, vzdelaním, národnosťou alebo náboženským vyznaním.

Z obsahovej stránky sú všetky využívané údaje definované v zmysle SODB 2021 [3].

Rodinný stav vyjadruje právny vzťah obyvateľov vzhľadom na manželstvo, ktorý platí v SR.

Pohlavie je definované ako biologická danosť človeka, môže byť mužské alebo ženské.

Vek sa v SODB zisťuje v dokončených rokoch na základe dátumu narodenia a rozhodujúceho okamihu sčítania. V analýze využívame 5-ročné vekové skupiny a hlavné vekové skupiny (0 – 14 rokov, 15 – 44 rokov, 45 – 64 rokov a 65 rokov a viac).

Región je definovaný na základe trvalého pobytu osoby. V analýze využívame údaje za kraje.

Počet živonarodených detí znamená počet všetkých detí, ktoré sa narodili obyvateľovi živé. V SODB 2021 sa tento údaj prvýkrát zisťoval aj u mužov. Počet živonarodených detí sa zisťoval len u 15-ročných a starších.

Pod vzdelaním sa rozumie najvyššie ukončené vzdelanie. Nie sú zahrnuté deti vo veku 0 – 14 rokov, ktoré vzhľadom na svoj vek nemôžu mať ukončené žiadne vzdelanie. Pod základným vzdelaním sa rozumie ukončený druhý stupeň základnej školy. Stredné vzdelanie bez maturity zahŕňa všetky osoby, ktoré absolvovali učňovské vzdelanie bez výučného listu, s výučným listom alebo s vysvedčením o záverečnej skúške. Stredné vzdelanie s maturitou majú všetky osoby, ktoré absolvovali učňovské, odborné alebo všeobecné vzdelanie ukončené maturitnou skúškou. Do tejto skupiny patria aj všetky formy vyššieho vzdelania. Do skupiny osôb s vysokoškolským vzdelaním patria všetky osoby, ktoré absolvovali vysokoškolské vzdelanie prvého, druhého alebo tretieho stupňa. Skupinu obyvateľov bez vzdelania tvoria osoby vo veku 15 rokov a viac bez ukončeného školského vzdelania.

Národnosť znamená príslušnosť obyvateľa k národu alebo etnickej skupine na základe seabedeklarácie. V analýze využívame pri hodnotení rodinného stavu obyvateľov päť najpočetnejších národností žijúcich na Slovensku – slovenskú, maďarskú, rómsku, rusínsku a českú. Zvyšné národnosti sú zahrnuté do spoločnej skupiny ostatné.

Náboženské vyznanie je členstvo alebo vzťah obyvateľa k cirkvi alebo náboženskej spoločnosti. Údaj je seabedeklaratórny. Pri hodnotení rodinného stavu obyvateľov využívame príslušnosť k štyrom najväčším cirkvám na Slovensku – rímskokatolíckej, evanjelickej augsburského vyznania, gréckokatolíckej a reformovanej kresťanskej cirkvi. Obyvatelia, ktorí sa prihlásili k niektorej zo zvyšných cirkví alebo náboženských spoločností, sú zahrnutí do skupiny ostatné. Skupina bez vyznania zahŕňa všetkých obyvateľov, ktorí sa neprihlásili ku žiadnej z cirkví alebo náboženských spoločností. Z hľadiska interpretácie je dôležité, že bez vyznania neznamená automaticky ateista, aj keď väčšinu obyvateľov v tejto skupine tvoria ateisti.

3. VÝVOJ OBYVATEĽOV SR PODĽA RODINNÉHO STAVU V OBDOBÍ 1991 – 2021

Keď hodnotíme celú populáciu Slovenska, dlhodobo prevládajú slobodné osoby a osoby žijúce v manželstve. Podiel obyvateľov v týchto dvoch skupinách presahuje dlhodobo 40 %. Podstatne nižšie zastúpenie majú v populácii na Slovensku rozvedené a ovdovené osoby, ktorých podiel je dlhodobo nižší ako 10 % (tabuľka č. 1, graf č. 1).

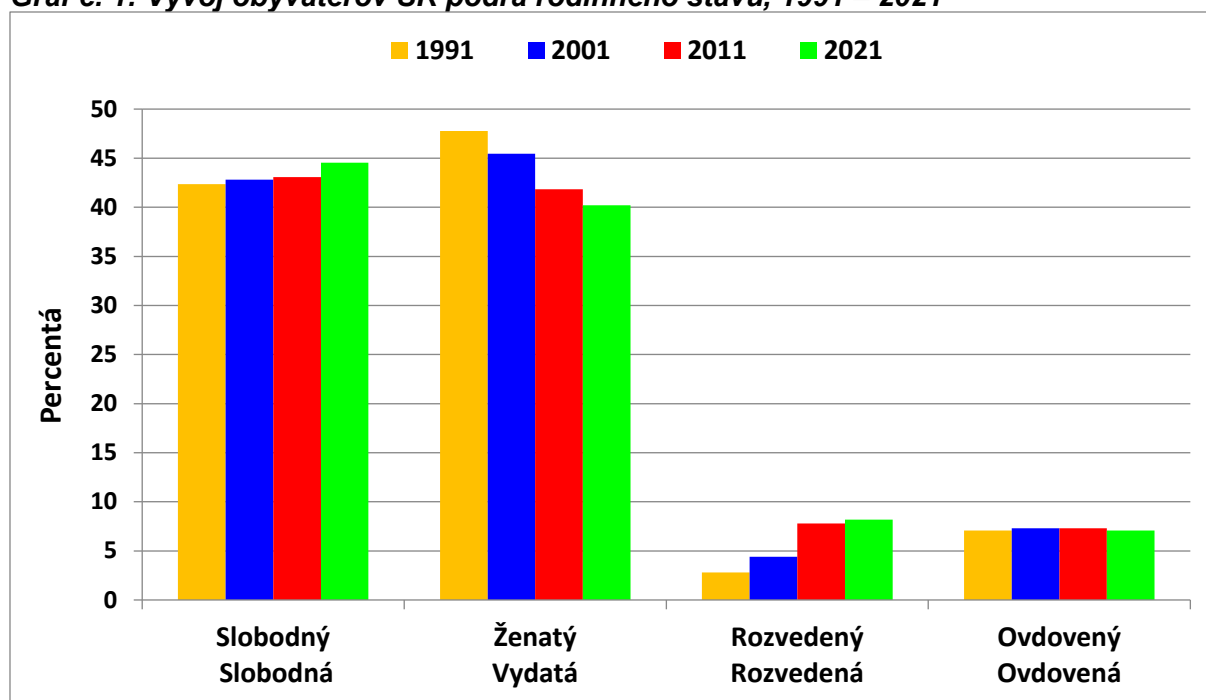
Hlavné trendy v štruktúre obyvateľov podľa rodinného stavu za posledné tri desaťročia možno v zásade vyjadriť ako mierny nárast podielu slobodných, výrazný pokles podielu ženatých, resp. vydatých, výrazné zvýšenie podielu rozvedených a stagnácia podielu ovdovených (tabuľka č. 1, graf č. 1). Ide o logický dôsledok aktuálneho demografického a spoločenského vývoja. Hlavným dôvodom uvedených zmien v štruktúre obyvateľov podľa rodinného stavu je klesajúca sobášnosť, rastúca rozvodovosť a zvyšovanie počtu kohabitácií [1, 2, 4]. Tieto trendy zvyšujú podiel slobodných, resp. rozvedených osôb na úkor osôb žijúcich v manželstve.

Tabuľka č. 1: Vývoj počtu obyvateľov SR podľa rodinného stavu, 1991 – 2021

Rok	Rodinný stav					Spolu
	slobodný slobodná	ženatý vydatá	rozvedený rozvedená	ovdovený ovdovená	nezistený	
1991	2 232 917	2 517 723	147 523	372 451	3 721	5 274 335
2001	2 273 779	2 414 523	233 552	388 392	69 209	5 379 455
2011	2 281 441	2 214 431	412 745	386 707	101 712	5 397 036
2021	2 418 756	2 182 365	444 536	383 625	19 988	5 449 270

Zdroj: Štatistický úrad SR

V období 1991 – 2011 sa podiel slobodných osôb v populácii Slovenska zvyšoval len mierne (intercenzálny nárast o menej ako 0,5 percentuálneho bodu). Výraznejší nárast podielu slobodných evidujeme v období 2011 – 2021 (zvýšenie o 1,5 percentuálneho bodu). Keď porovnáme začiatok a koniec hodnoteného obdobia, podiel slobodných osôb sa v populácii Slovenska zvýšil o 2,2 percentuálneho bodu, čo znamená priemerný intercenzálny nárast o 0,73 percentuálneho bodu.

Graf č. 1: Vývoj obyvateľov SR podľa rodinného stavu, 1991 – 2021**Zdroj: Štatistický úrad SR**

Podiel ženatých, resp. vydatých osôb v populácii Slovenska dlhodobo klesá. Počas sledovaného obdobia sa podiel osôb žijúcich v manželstve znížil o viac ako 7,5 percentuálneho bodu, čo znamená priemerný intercenzálny pokles o 2,52 percentuálneho bodu. Podiel ženatých, resp. vydatých osôb sa v populácii na Slovensku znižoval intenzívnejšie v období 1991 – 2011, aj v období 2011 – 2021 bol však pokles zrejмый.

Podiel rozvedených osôb sa v populácii na Slovensku zvýšil počas hodnoteného obdobia o 5,4 percentuálneho bodu. Vzhľadom na početnosť rozvedených osôb v populácii ide o najväčšiu zmenu zo všetkých rodinných stavov (podiel rozvedených

osôb sa v populácii počas obdobia 1991 – 2021 zvýšil viac ako trojnásobne). Najväčší nárast počtu rozvedených mužov a žien evidujeme v období 2001 – 2011 (o 3,4 percentuálneho bodu), výrazné zvýšenie nastalo aj v období 1991 – 2001 (o 1,6 percentuálneho bodu). V poslednom intercenzálnom období sa podiel rozvedených zvýšil len mierne (zhruba o 0,4 percentuálneho bodu). V priemere dosahoval intercenzálny nárast rozvedených osôb v populácii Slovenska 1,8 percentuálneho bodu.

Najmenšiu zmenu počas analyzovaného obdobia zaznamenali ovdovené osoby. Intercenzálne zmeny podielu ovdovených osôb v populácii sa pohybovali pod hranicou 0,25 percentuálneho bodu, a keď porovnáme začiatok a koniec hodnoteného obdobia, tak podiel ovdovených v populácii zostal nezmenený, pričom tesne presiahol hranicu 7 %.

4. OBYVATEĽSTVO PODĽA RODINNÉHO STAVU V KRAJOCH SR

Regionálne rozdiely na Slovensku sú vo všeobecnosti veľmi výrazné a platí to aj pre reprodukčné správanie obyvateľov [1]. Preto je zrejmé, že viditeľné regionálne rozdiely existujú, aj čo sa týka štruktúry obyvateľov podľa rodinného stavu. Pri hodnotení regionálnych rozdielov v rodinnom stave obyvateľov berieme do úvahy celú populáciu a hodnotíme ju na úrovni krajov¹.

Tabuľka č. 2: Rodinný stav obyvateľov v krajoch SR v roku 2021

Kraj	Rodinný stav					Spolu
	slobodný slobodná	ženaný vydatá	rozvedený rozvedená	ovdovený ovdovená	nezistený	
Bratislavský	319 571	281 411	68 877	41 976	7 702	719 537
Trnavský	242 429	230 151	49 791	41 025	2 612	566 008
Trenčiansky	244 051	239 097	50 706	42 189	1 421	577 464
Nitriansky	289 917	266 555	63 597	55 044	2 787	677 900
Žilinský	307 131	285 568	49 670	48 070	1 174	691 613
Banskobystrický	282 460	233 297	59 584	48 989	1 271	625 601
Prešovský	365 488	346 483	44 237	51 730	993	808 931
Košický	367 709	299 803	58 074	54 602	2 028	782 216

Zdroj: Štatistický úrad SR

Vo všetkých krajoch SR je zachovaná rovnaká základná štruktúra obyvateľov podľa rodinného stavu ako na celoslovenskej úrovni. Najviac obyvateľov je slobodných, nasledujú osoby žijúce v manželstve, ďalej rozvedení a najmenej je ovdovených. Jedinou výnimkou je Prešovský kraj s nízkou rozvodovosťou, v dôsledku čoho je v tomto kraji menej rozvedených ako ovdovených obyvateľov (tabuľka č. 2, graf č. 2).

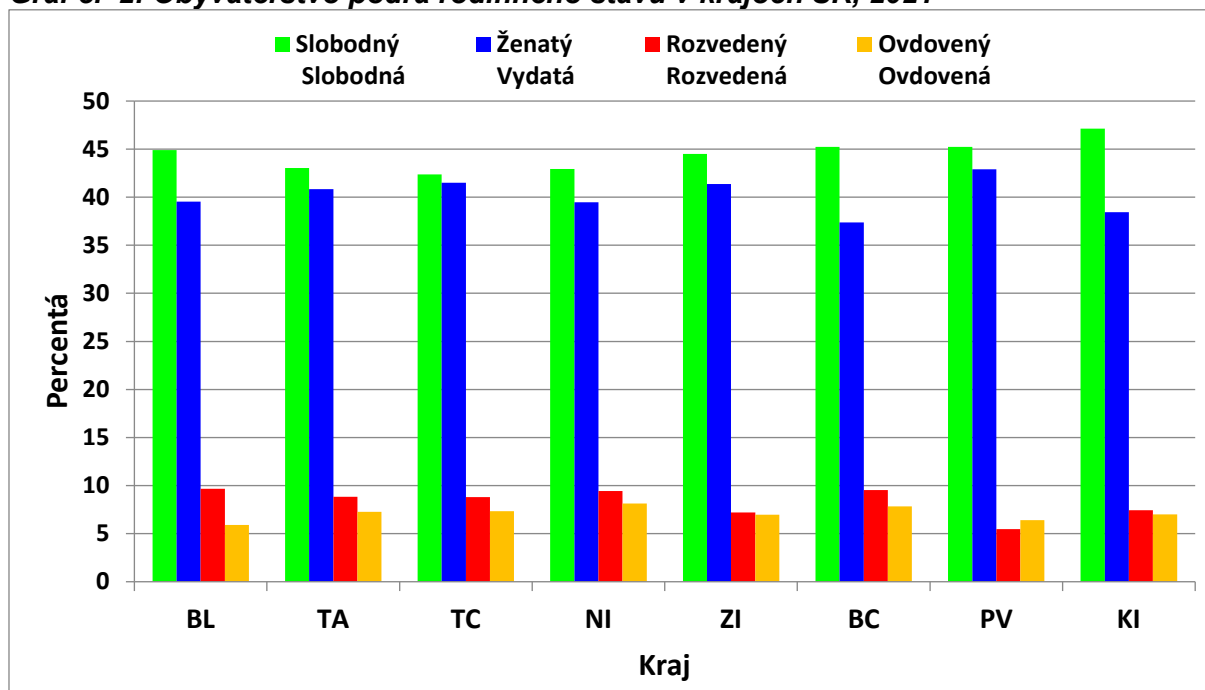
Aj keď základné trendy sú na úrovni krajov rovnaké, regionálne rozdiely sa prejavujú v detailnejšom pohľade na rodinný stav obyvateľov.

Podiel slobodných osôb presiahol v roku 2021 vo všetkých krajoch 42 %. Najnižší bol v Trnavskom kraji, (42,8 %), pod úroveň 44 % bol podiel slobodných aj v Trenčianskom a Nitrianskom kraji. Na opačnom konci sú kraje s podielom

¹ Podrobnejšia analýza (napr. na úrovni okresov) by lepšie prezentovala regionálne rozdiely v rodinnom stave obyvateľstva, rozsahom však presahuje možnosti tohto článku.

slobodných vyšším ako 45 %. Ide o Banskobystrický, Prešovský a Košický kraj, v ktorom je najvyšší podiel slobodných osôb 47,1 %. Vo všetkých týchto troch krajoch je na slovenské pomery vysoká pôrodnosť a tým aj vyššie zastúpenie detí v populácii, čo sa logicky prejavuje aj na vyššom podiele slobodných osôb v populácii.

Graf č. 2: Obyvateľstvo podľa rodinného stavu v krajoch SR, 2021



Zdroj: Štatistický úrad SR

Podiel ženatých, resp. vydatých osôb sa v roku 2021 pohyboval v jednotlivých krajoch v rozpätí od 37,4 % v Banskobystrickom kraji po 42,9 % v Prešovskom kraji. Podpriemerné zastúpenie osôb žijúcich v manželstve bolo v Košickom kraji (38,4 %), nadpriemerné v Trenčianskom a Žilinskom kraji (41,5 %, resp. 41,4 %).

Podiel rozvedených sa v roku 2021 pohyboval v jednotlivých krajoch od 5 % do 10 %, pričom kopíroval regionálne rozdiely v rozvodovosti. Najnižší podiel rozvedených osôb bol v Prešovskom kraji (5,5 %), relatívne nízky podiel rozvedených bol aj v Žilinskom a Košickom kraji (7,2 %, resp. 7,4 %). Najviac rozvedených (viac ako 9,5 %) bolo v Bratislavskom a Banskobystrickom kraji.

Regionálne rozdiely v počte ovdovených sú najmenšie zo všetkých rodinných stavov. Podiel ovdovených v roku 2021 sa v jednotlivých krajoch pohyboval v úzkom rozpätí od 5,9 % v Bratislavskom kraji po 8,1 % v Nitrianskom kraji. Medzi kraje s nižším podielom ovdovených osôb patrili Žilinský a Prešovský kraj (podiel ovdovených menej ako 7 %), nadpriemerný podiel ovdovených osôb bol v Banskobystrickom kraji (9,5 %).

5. ZÁKLADNÉ ŠTRUKTÚRY RODINNÉHO STAVU OBYVATEĽOV SR

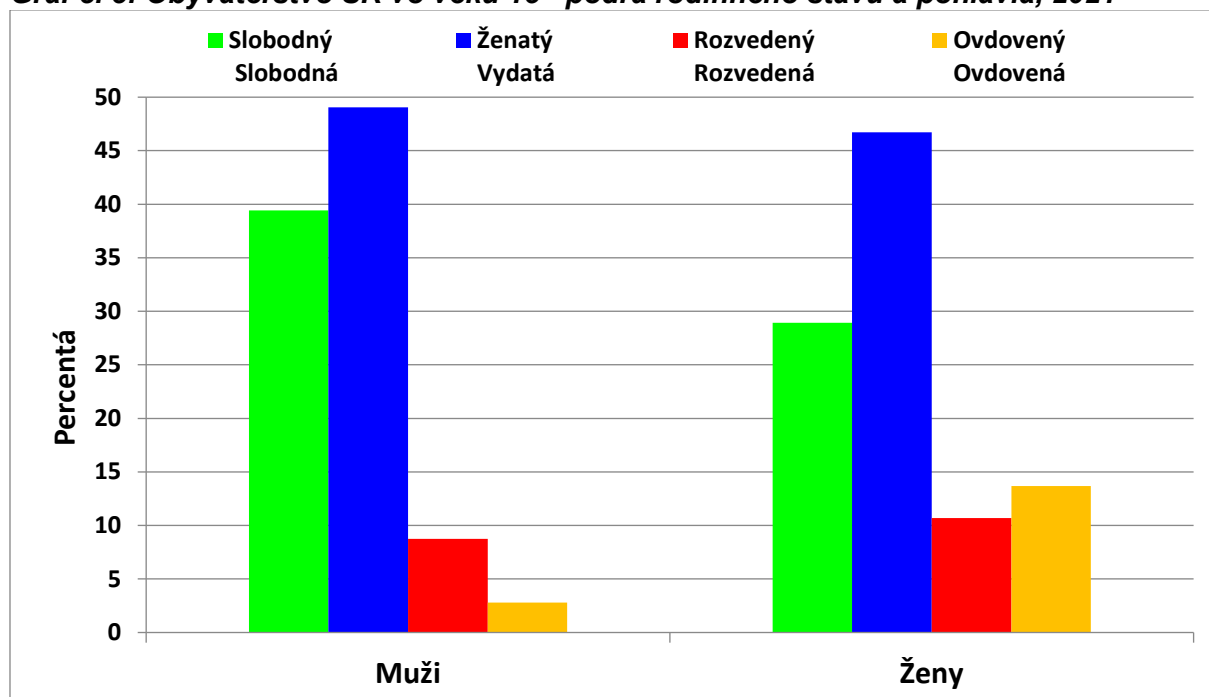
Na reprodukčné správanie obyvateľstva vplyvajú mnohé faktory a tie spôsobujú aj rozdiely v štruktúre obyvateľov podľa rodinného stavu. Ako už bolo spomenuté v úvode, do hodnotenia štruktúry obyvateľov podľa rodinného stavu sme zvolili tie faktory, ktoré významne ovplyvňujú reprodukčné správanie obyvateľov a zároveň sú k dispozícii v SODB 2021.

5.1. POHLAVIE

Reprodukčné správanie obyvateľov sa líši podľa pohlavia, preto sa v závislosti od pohlavia líši aj štruktúra obyvateľov podľa rodinného stavu [1, 2, 4].

Najmarkantnejší je vplyv nadúmrtosti mužov, ktorý spôsobuje, že počet aj podiel ovdovených žien v populácii Slovenska je podstatne vyšší, ako počet a podiel ovdovených mužov. Príčinou vyššieho podielu rozvedených žien v populácii je zas častejšia opakovaná sobášnosť mužov. V populácii Slovenska je dlhodobo výrazne vyšší podiel slobodných mužov ako slobodných žien. Príčinou je hlavne to, že muži vstupujú do manželstva v priemere vo vyššom veku ako aj to, že častejšie neuzatvárajú manželstvo vôbec a zostávajú slobodní. Počet mužov a žien žijúcich v manželstve je v populácii vyrovnaný, rozdiel medzi zisteným počtom ženatých a vydatých osôb v populácii spôsobujú osoby s nezisteným rodinným stavom (tabuľka č. 3, graf č. 3).

Graf č. 3: Obyvateľstvo SR vo veku 15+ podľa rodinného stavu a pohlavia, 2021



Zdroj: Štatistický úrad SR

U mužov vo veku 15+ tvoria dlhodobo najvyšší podiel ženatí muži, v roku 2021 to bolo takmer 50 % z celkového počtu. Nasledovali slobodní muži, ktorých podiel na celkovom počte v roku 2021 bol o 10 percentuálnych bodov nižší. Rozvedení muži tvorili v rovnakom období 8,7 % z celkového počtu a najmenší podiel na celkovom počte mužov vo veku 15+ evidujeme u ovdovelých mužov (zhruba 2,8 %) (tabuľka č. 3, graf č. 3).

Situácia u žien je odlišná. V roku 2021 bol podiel vydatých aj slobodných žien v populácii 15+ nižší ako u mužov a rozdiel medzi podielom slobodných a vydatých žien bol väčší ako v prípade mužov (u žien to bolo zhruba 15 percentuálnych bodov a u mužov len 10). Ovdovených žien bolo viac ako rozvedených (13,7 % vs. 10,7 %), čo je opačná situácia ako u mužov. Obidva tieto rodinné stavy sú v populácii žien zastúpené vyšším podielom ako v populácii mužov (tabuľka č. 3, graf č. 3).

Tabuľka č. 3: Obyvateľstvo SR vo veku 15+ podľa rodinného stavu v roku 2021

Rok	Rodinný stav					Spolu
	slobodný slobodná	ženatý vydatá	rozvedený rozvedená	ovdovený ovdovená	nezistený	
Pohlavie						
Muži	870 248	1 082 436	192 925	61 773	13 718	2 221 100
Ženy	681 098	1 099 929	251 611	321 852	6 270	2 360 760
Vekové skupiny						
15-44	1 305 511	750 559	104 280	5 420	11 390	2 177 160
45-64	197 912	939 467	257 598	73 815	6 727	1 475 519
65+	47 923	492 339	82 658	304 390	1 871	929 181
Počet detí						
0	1 132 853	131 838	26 751	12 289	84	1 303 815
1	192 481	394 544	121 759	42 668	2 096	753 548
2	81 115	1 053 055	192 242	160 457	1 361	1 488 230
3	23 558	410 270	70 693	101 359	347	606 227
4	10 350	115 695	19 698	38 512	76	184 331
5+	14 254	67 291	9 518	26 580	43	117 686
nezistený	96 735	9 672	3 875	1 760	15 981	128 023
Vzdelanie						
Základné	571 844	195 553	38 491	118 510	210	924 608
Stredné bez mat.	241 918	568 983	123 233	113 194	25	1 047 353
Stredné s mat.	539 268	795 512	169 199	107 649	45	1 611 673
Vysokoškolské	325 139	562 553	83 159	30 091	504	1 001 446
Bez vzdelania	11 546	2 145	443	851	2	14 987
Nezistené	105 842	57 619	30 011	13 330	19 202	226 004
Národnosť						
Slovenská	1 297 689	1 877 890	368 024	316 711	141	3 860 455
Maďarská	113 451	177 715	38 782	42 448	325	372 721
Rómska	21 758	18 403	1 403	1 861	9	43 434
Rusínska	5 279	12 296	1 399	2 818	2	21 794
Česká	4 271	15 948	3 806	3 858	99	27 982
Ostatné	9 992	23 803	3 631	2 489	447	40 362
Nezistená	98 906	56 310	27 491	13 440	18 965	215 112
Náboženské vyznanie						
Rímskokatolícke	761 604	1 327 197	222 472	252 294	200	2 563 767
Evanjelické a. v.	70 207	127 770	24 870	30 106	7	252 960
Gréckokatolícke	53 417	96 995	13 474	17 519	8	181 413
Reformovaní	22 843	36 438	6 851	8 708	5	74 845
Ostatné	55 823	70 619	12 211	8 766	54	147 473
Bez vyznania	464 198	454 683	129 140	50 891	157	1 099 069
Nezistené	123 254	68 663	35 518	15 341	19 557	262 333

Zdroj: Štatistický úrad SR

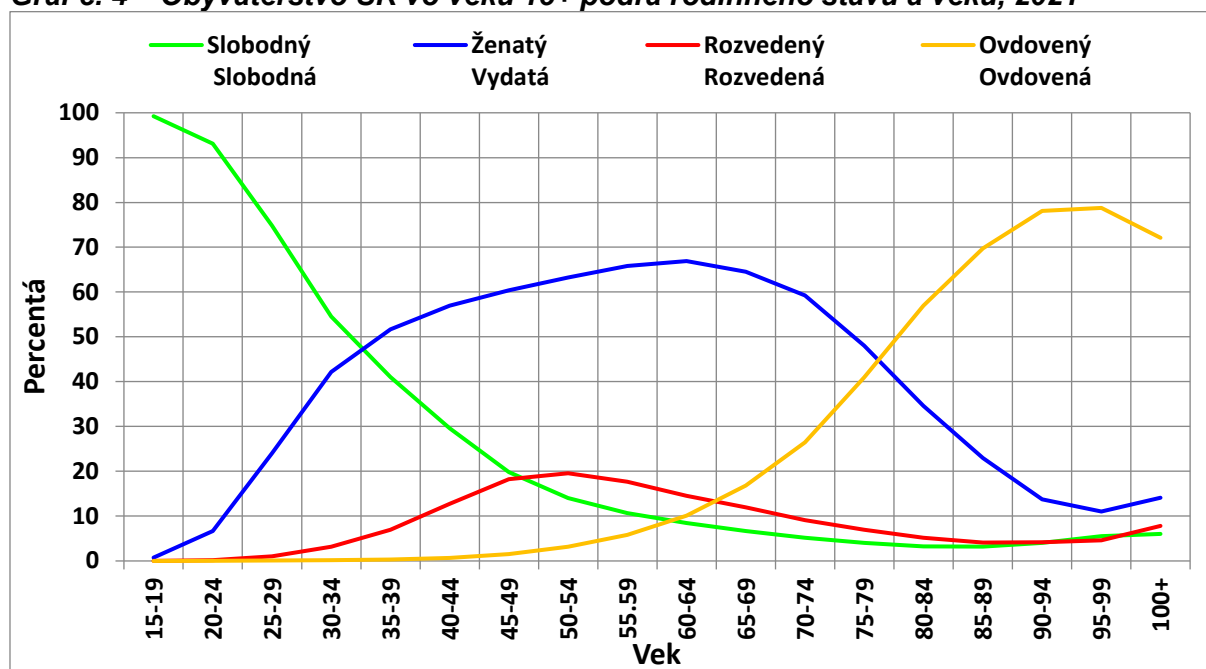
5.2. VEK

Podobne ako pohlavie aj vek má výrazný vplyv na reprodukčné správanie obyvateľov [1, 2, 4]. Preto aj rozdiely v rodinnom stave obyvateľov v závislosti od veku sú výrazné.

V detskom veku sú všetci obyvatelia slobodní. Stopercentný podiel slobodných v populácii sa končí vo veku 16 rokov, keď je možné (aj keď len za špecifických podmienok) vstupovať do manželstva. Ide však naozaj len o výnimočné prípady, preto reálny počet ženatých, resp. vydatých osôb sa začína zvyšovať až od veku 18 rokov, kedy je možné vstupovať do manželstva bez súdneho rozhodnutia. Pretože priemerný vek pri prvom sobáši sa u oboch pohlaví neustále zvyšuje, podiel slobodných osôb v mladom veku sa stále dlhšie drží nad hranicou 90 %. V roku 2021 to bolo až do veku 23 rokov. Od tohto veku sa podiel slobodných v populácii Slovenska prudko znižoval, pričom výrazný pokles trval až do veku 55 rokov. Do veku 35 rokov však bolo slobodné obyvateľstvo najpočetnejšou skupinou zo všetkých rodinných stavov. Od veku 55 rokov sa podiel slobodných obyvateľov v populácii znižoval už len mierne a zhruba od 60. roku života sa slobodní stali najmenej početnou skupinou obyvateľov s podielom menej ako 10 % z celkového počtu obyvateľov (tabuľka č. 3, graf č. 4).

Prudký nárast podielu ženatých a vydatých osôb v populácii, ktorý sa začína vo veku 20 rokov, vrcholil v roku 2021 vo veku 60 – 64 rokov, keď podiel osôb žijúcich v manželstve dosiahol takmer 70 % z celkového počtu obyvateľov. Od veku 65 rokov sa podiel ženatých a vydatých osôb v populácii postupne znižoval, a to až k hranici 10 % vo veku nad 90 rokov. Vo veku 35 – 77 rokov boli osoby žijúce v manželstve najpočetnejšou skupinou v populácii Slovenska zo všetkých rodinných stavov (tabuľka č. 3, graf č. 4).

Graf č. 4 – Obyvateľstvo SR vo veku 15+ podľa rodinného stavu a veku, 2021



Zdroj: Štatistický úrad SR

Podiel rozvedených obyvateľov sa začína zvyšovať od veku 25 rokov. V roku 2021 dosiahol podiel rozvedených vrchol vo veku 50 – 54 rokov a to zhruba 20 %

z celkového počtu obyvateľov. Následný pokles znamenal zníženie podielu rozvedených osôb v populácii Slovenska až ku hranici 5 % vo veku nad 80 rokov (tabuľka č. 3, graf č. 4).

Podiel ovdovených sa začína zvyšovať po dosiahnutí veku 50 rokov. V roku 2021 boli ovdovení najpočetnejšou skupinou zo všetkých rodinných stavov v populácii Slovenska od 78 rokov, pričom ich podiel sa zvyšoval až do 95 rokov, keď až 80 % obyvateľov Slovenska patrilo do skupiny ovdovených.

5.3. POČET ŽIVONARODENÝCH DETÍ

Počet narodených detí úzko súvisí s vekom a trvaním manželstva, resp. partnerského spolužitia. S narastajúcim počtom detí sa zvyšuje vek rodičov, ubúda slobodných a pribúda ženatých, resp. vydatých a ovdovených. Rozvedení aspoň s jedným dieťaťom si udržiavajú pomerne stabilný podiel v populácii Slovenska (tabuľka č. 3, graf č. 5).

V skupine bezdetných obyvateľov bolo v roku 2021 takmer 90 % slobodných. Ide nielen o osoby, ktoré nemajú deti preto, že z rôznych dôvodov žijú bez partnera (dočasne alebo trvale) ale tiež o bezdetné kohabitujúce páry. V skupine osôb s jedným dieťaťom bol podiel slobodných ešte pomerne vysoký (zhruba 25 %). Vyplýva to z povahy kohabitácií, keď kohabitujúce páry majú vo všeobecnosti menej detí ako manželské páry a to aj preto, že po narodení detí sa kohabitácie často menia na manželstvo [1, 2, 4]. Podiel slobodných osôb v skupine obyvateľov s viac ako jedným dieťaťom bol v roku 2021 už výrazne nižší. Pri dvoch, troch a štyroch deťoch bol podiel slobodných osôb menší ako 6 %, v skupine obyvateľov s 5 a viac deťmi sa zvýšil nad hranicu 10 %. Vyšší podiel slobodných v tejto skupine obyvateľov je spôsobený vyšším podielom Rómov medzi mnohodetnými obyvateľmi. Je známe, že Rómovia majú výrazne vyššie zastúpenie slobodných vo svojej populácii ako majoritné obyvateľstvo [5].

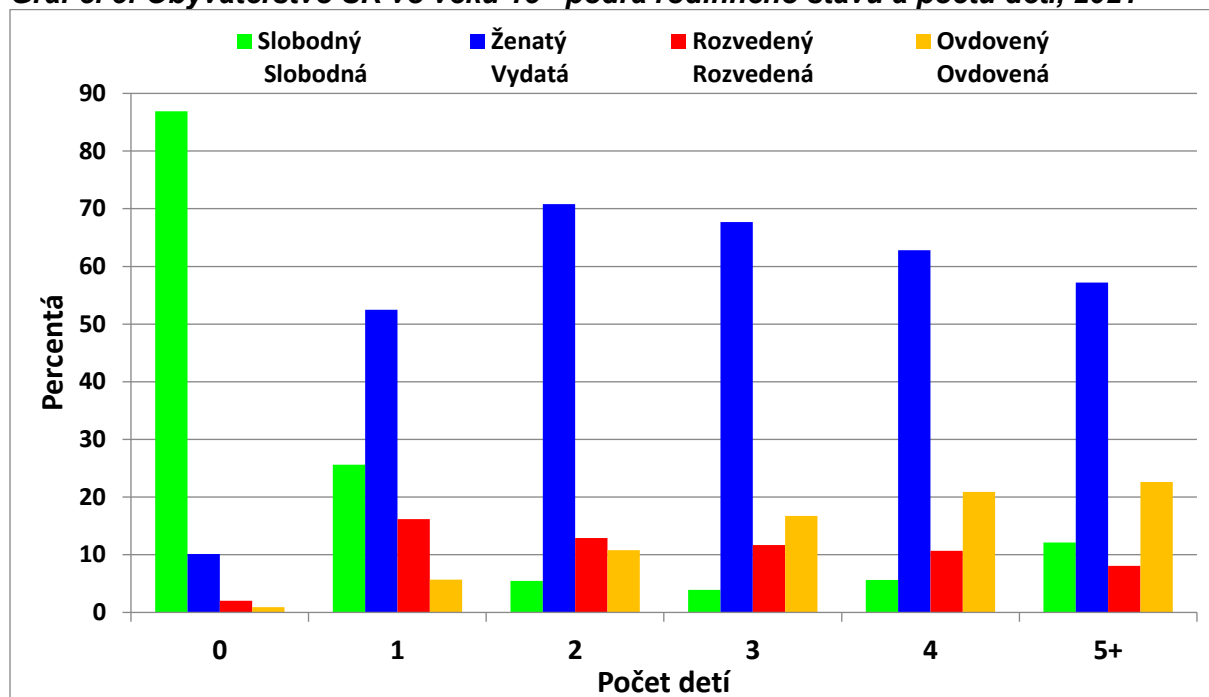
Osoby žijúce v manželstve tvorili v roku 2021 v skupine bezdetných obyvateľov len 10 %, čiastočne ide o manželské páry, ktoré nemôžu mať deti a čiastočne o ženatých, resp. vydaté, ktorí nechcú mať deti, pričom môže ísť o trvalú životnú stratégiu alebo o odklad narodenia dieťaťa do vyššieho veku [1, 2, 4]. Pri obyvateľoch aspoň s jedným dieťaťom dominujú ženaté, resp. vydaté osoby. V prípade obyvateľov s dvomi, tromi alebo štyrmi deťmi sa podiel osôb žijúcich v manželstve pohyboval v roku 2021 od 60 do 70 %. Menší podiel (od 50 % do 60 %) tvorili ženaté, resp. vydaté osoby v skupine obyvateľov s jedným dieťaťom a s 5 a viac deťmi. V skupine jednodetných rodín znižujú podiel ženatých, resp. vydatých partnerské spolužitia, u mnohodetných rodín zas vzhľadom na vyšší vek manželov častejšie evidujeme ovdovenie jedného z nich.

Najväčší podiel rozvedených osôb je v skupine obyvateľov s jedným dieťaťom (v roku 2021 to bolo 16,2 %). U obyvateľov najmenej s dvomi deťmi sa podiel rozvedených osôb pohyboval v blízkosti hranice 10 % s mierne klesajúcou tendenciou pri rastúcom počte detí. Tento stav zodpovedá časovaniu rozvodov a tiež skutočnosti, že rozvodovosť sa znižuje so zvyšujúcim sa počtom detí. Na Slovensku dlhodobo zhruba $\frac{3}{4}$ rozvodov tvoria manželstvá bez detí alebo s jedným dieťaťom [1, 2, 4].

So zvyšujúcim sa počtom detí sa zvyšuje podiel ovdovených osôb. Dá sa predpokladať, že väčší počet detí v rodine je spojený s vyšším vekom manželov a teda

aj s väčšou pravdepodobnosťou, že jeden z manželov ovdovie. V skupine obyvateľov s menej ako 2 deťmi bol podiel ovdovených v roku 2021 nízky, nedosahoval ani 6 %. V skupine obyvateľov s dvomi deťmi sa pohyboval tesne nad hranicou 10 % a postupne sa zvyšoval na viac ako 22 % ovdovených v skupine obyvateľov s 5 a viac deťmi.

Graf č. 5: Obyvateľstvo SR vo veku 15+ podľa rodinného stavu a počtu detí, 2021



Zdroj: Štatistický úrad SR

5.4. VZDELANIE

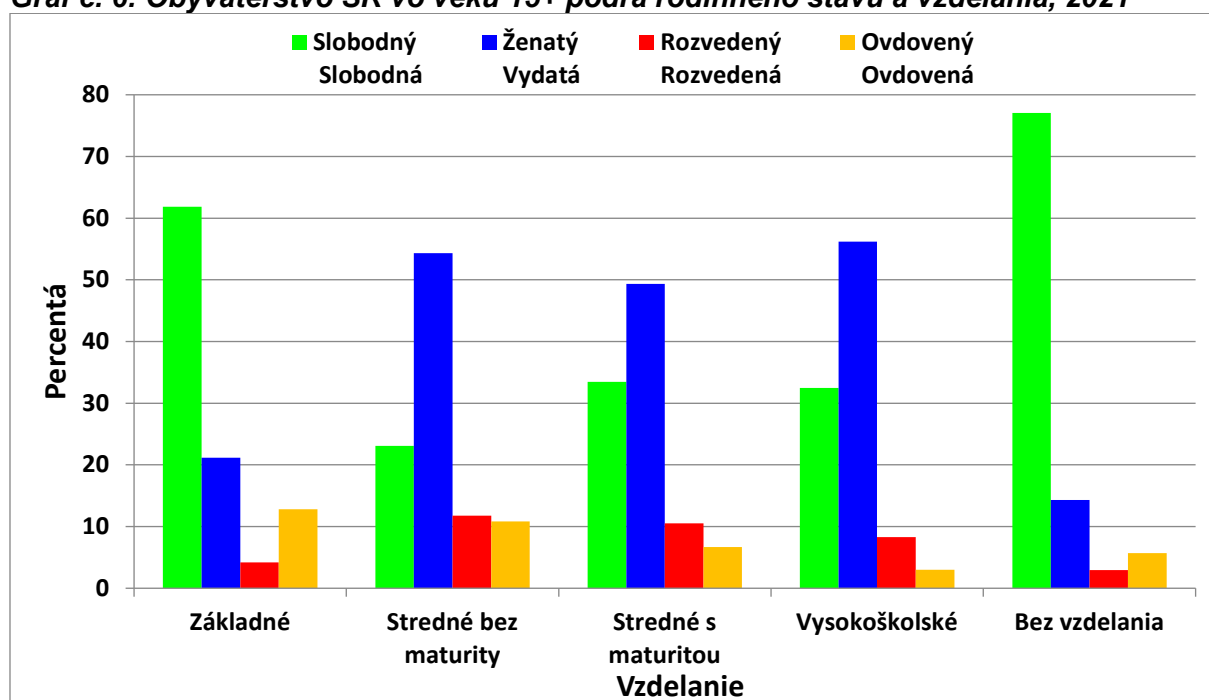
V skupine obyvateľov so základným vzdelaním a bez vzdelania výrazne prevládajú slobodné osoby (v roku 2021 viac ako 60 %, resp. takmer 80 %) – tabuľka č.3, graf č. 6. Jedným z dôvodov vysokého podielu slobodných v skupine obyvateľov so základným vzdelaním je aj skutočnosť, že do tejto skupiny patria osoby študujúce na stredných školách (s neukončeným stredoškolským vzdelaním), ktoré vzhľadom na svoj mladý vek vstupujú do manželstva len výnimočne. V prípade osôb bez školského vzdelania sa zas možno domnievať, že do tejto skupiny obyvateľov patria vo väčšej miere osoby rôznym spôsobom hendikepované, ktoré počas svojho života neuzatvárajú manželstvo a nezakladajú si rodiny. Najmenej slobodných osôb zo všetkých vzdelanostných skupín obyvateľov je medzi obyvateľmi so stredným vzdelaním bez maturity v roku 2021 to bolo 23 %. V skupine stredoškolákov s maturitou a vysokoškolákov presahoval podiel slobodných obyvateľov 30 %.

V skupine obyvateľov so stredným a vysokoškolským vzdelaním prevládajú osoby žijúce v manželstve. Prvoradým dôvodom je vek, keďže ide o osoby, ktoré už môžu vstupovať do manželstva bez vekového obmedzenia. V roku 2021 sa podiel ženatých alebo vydatých osôb v skupine obyvateľov so stredoškolským alebo vysokoškolským vzdelaním pohyboval od 50 do 60 %. Osoby so základným vzdelaním tvorili viac ako 20 % a osoby bez vzdelania len 14 % z celkového počtu obyvateľov v týchto vzdelanostných skupinách.

Nízky podiel rozvedených v skupine obyvateľov so základným vzdelaním a bez vzdelania súvisí hlavne s nižším počtom manželstiev, ktoré obyvateľstvo v týchto dvoch vzdelanostných skupinách uzatvára. V ostatných troch vzdelanostných skupinách (stredné vzdelanie s maturitou aj bez maturity a vysokoškolské vzdelanie) sa podiel rozvedených pohyboval v roku 2021 blízko hranice 10 %.

Je známe, že so zvyšujúcim sa vzdelaním sa znižuje úmrtnosť [1, 2, 4]. So zvyšujúcim sa vzdelaním preto logicky klesá aj podiel ovdovených osôb, ktorý priamo súvisí s predčasnými úmrtiami, ktorých je u obyvateľov s vyšším vzdelaním menej. Najvyšší podiel ovdovených osôb bol v roku 2021 v skupine obyvateľov so základným vzdelaním (zhruba 13 %) a najnižší v skupine obyvateľov s vysokoškolským vzdelaním (3 %).

Graf č. 6: Obyvateľstvo SR vo veku 15+ podľa rodinného stavu a vzdelania, 2021



Zdroj: Štatistický úrad SR

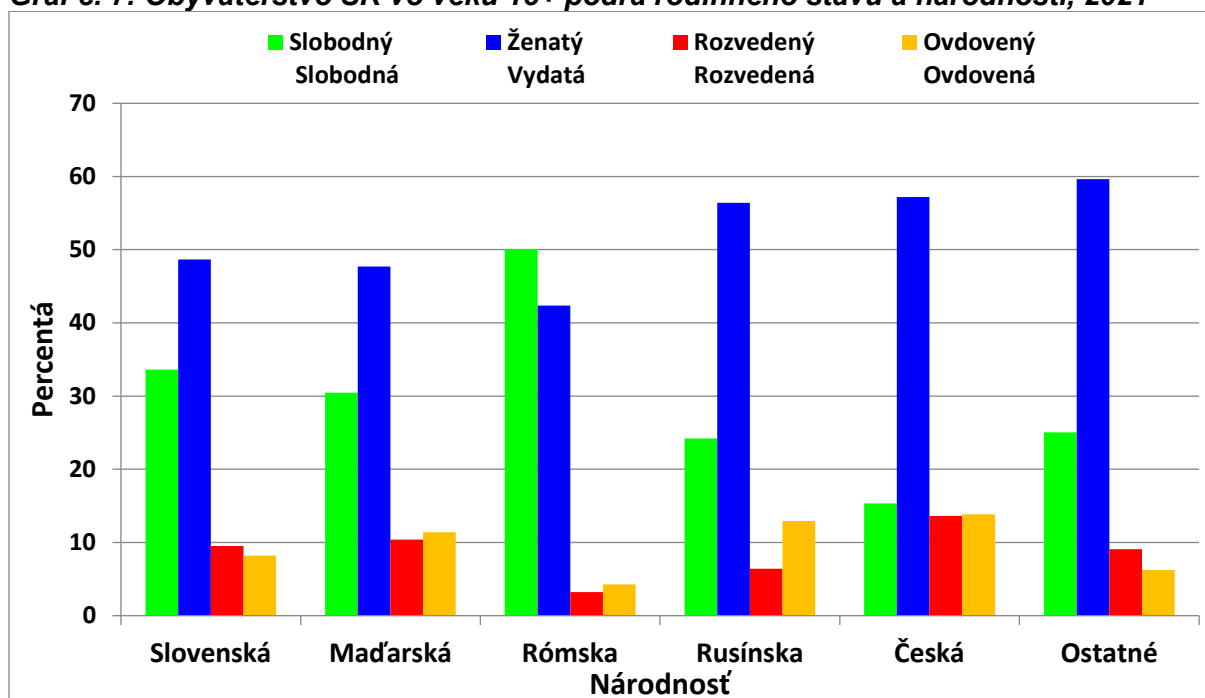
5.5. NÁRODNOSŤ

Národnosť v sebe spája kultúrne, regionálne aj historické špecifiká, ktoré ovplyvňujú reprodukčné správanie obyvateľstva a tým aj štruktúru obyvateľov podľa rodinného stavu.

Podobná štruktúra obyvateľov podľa rodinného stavu je u obyvateľov slovenskej a maďarskej národnosti s pomerne malou prevahou osôb žijúcich v manželstve nad slobodnými osobami (v roku 2021 bol tento rozdiel necelých 20 percentuálnych bodov v prospech ženatých, resp. vydatých osôb). V obidvoch populáciách bol podobný aj podiel rozvedených a ovdovených osôb (v roku 2021 menej ako 10 %). Zatiaľ čo u obyvateľov slovenskej národnosti mierne prevládali rozvedení nad ovdovenými, u obyvateľov maďarskej národnosti to bolo naopak (tabuľka č. 3, graf č. 7).

U obyvateľov rómskej národnosti dlhodobo výrazne prevládajú slobodné osoby, ich podiel v populácii 15+ bol v roku 2021 50 %. Dôvodom je vysoká pôrodnosť a tým vysoký počet detí v populácii, vysoká úmrtnosť, ktorá znižuje počet seniorov v populácii a tým aj počet rozvedených osôb a tiež nižšia sobášnosť, keď oficiálne sobáše sú čiastočne nahrádzané rómskymi svadbami, ktoré nemajú vplyv na rodinný stav obyvateľov [5]. Podiel osôb žijúcich v manželstve je u obyvateľov rómskej národnosti najnižší zo všetkých hodnotených národnostných skupín (v roku 2021 bol tesne nad hranicou 40 %). Podiel rozvedených a ovdovených je u obyvateľov rómskej národnosti veľmi nízky (3,2 %, resp. 4,3 % v roku 2021). Je známe, že Rómovia majú nízku rozvodovosť a vysokú úmrtnosť mužov aj žien, v dôsledku čoho neevidujeme tak vysokú nadúmrtnosť mužov, ktorá je hlavným dôvodom vyššieho podielu ovdovených u nerómskej populácie (tabuľka č. 3, graf č. 7).

Graf č. 7: Obyvateľstvo SR vo veku 15+ podľa rodinného stavu a národnosti, 2021



Zdroj: Štatistický úrad SR

U obyvateľov rusínskej a českej národnosti ako aj v skupine ostatných národností výrazne prevládajú osoby žijúce v manželstve s podielom na celkovom počte obyvateľov od 55 % do 60 %. Druhou najpočetnejšou skupinou u týchto troch národností sú slobodné osoby. U obyvateľov rusínskej národnosti a v skupine ostatných národností dosahoval podiel slobodných osôb v roku 2021 zhruba 25 %, u obyvateľov českej národnosti len 15 %. Podiel rozvedených a ovdovených osôb sa u obyvateľov rusínskej, českej a ostatných národností pohyboval v roku 2021 od 6 % do 14 %. U obyvateľov českej národnosti bol podiel rozvedených a ovdovených vyrovnaný a najvyšší zo všetkých hodnotených národností (viac ako 13 % v roku 2021). Obyvateľstvo rusínskej národnosti (s nízkou rozvodovosťou) malo v roku 2021 druhý najnižší podiel rozvedených osôb po rómskej národnosti (6,4 %) a pomerne vysoký podiel ovdovených osôb (takmer 13 %). V skupine ostatných národností bol v roku 2021 podiel rozvedených aj ovdovených obyvateľov nižší ako 10 %, pričom menej bolo ovdovených osôb ako rozvedených.(tabuľka č. 3, graf č. 7).

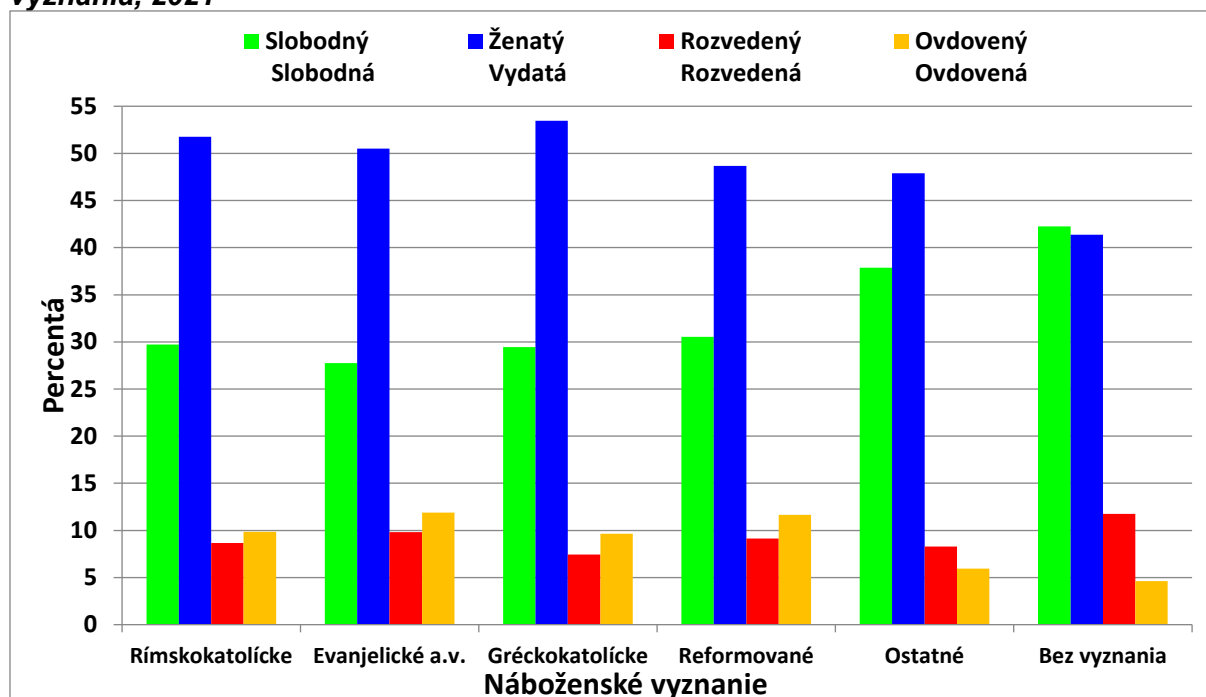
5.6. VIEROVYZNANIE

Pri štyroch najpočetnejších náboženských vierovyznaniach (ide o kresťanské cirkvi – rímskokatolícku, evanjelickú augsburského vyznania, gréckokatolícku a reformovanú) nie sú výrazné rozdiely v rozdelení obyvateľov podľa rodinného stavu.

Výrazne najpočetnejšiu skupinu tvoria obyvatelia žijúci v manželstve, ich podiel sa v roku 2021 pohyboval v blízkosti hranice 50 % (u reformovaných tesne pod touto hranicou pri zvyšných troch vierovyznaniach tesne nad ňou). Nasledovali slobodné osoby s podielom na úrovni 30 % (len u evanjelikov a. v. bol podiel zhruba o 2 percentuálne body nižší). Podiel rozvedených a ovdovených obyvateľov sa pohyboval v blízkosti hranice 10 %, pričom pri všetkých štyroch náboženských vierovyznaniach bol v roku 2021 vyšší podiel ovdovených ako rozvedených. Vyšší podiel rozvedených aj ovdovených bol u evanjelikov a. v. a reformovaných, nižší u katolíkov (západného aj východného obradu) (tabuľka č. 3, graf č. 8).

Výrazne odlišná je štruktúra obyvateľov podľa rodinného stavu v skupine ostatných cirkví a hlavne v skupine obyvateľov bez náboženského vyznania. V oboch týchto skupinách obyvateľov je vysoký podiel slobodných osôb a nižší podiel osôb žijúcich v manželstve (u obyvateľov bez vyznania sú slobodné osoby dokonca najpočetnejšou skupinou v populácii 15+). U rozvedených a ovdovených je situácia opačná ako evidujeme v štyroch najväčších cirkvách. V skupine ostatných cirkví a u obyvateľov bez náboženského vyznania je vyšší podiel rozvedených osôb ako ovdovených (tabuľka č. 3, graf č. 8).

Graf č. 8 – Obyvateľstvo SR vo veku 15+ podľa rodinného stavu a náboženského vyznania, 2021



Zdroj: Štatistický úrad SR

Aj uvedené skutočnosti potvrdzujú, že náboženské vierovyznanie ovplyvňuje reprodukčné správanie obyvateľov. Rozdiely sa týkajú všetkých reprodukčných procesov azda s výnimkou úmrtnosti [1, 2, 4]. Viac slobodných a menej ženatých, resp. vydatých súvisí s častejšími kohabitáciami v skupine obyvateľov bez vyznania, viac

rozvedených s vyššou rozvodovosťou v tejto skupine obyvateľov a menej ovdovených s nárastom religiozity vo vyššom veku.

6. ZÁVER

Rodinný stav je dôležitou charakteristikou obyvateľstva. Poskytuje obraz o obyvateľstve z hľadiska reprodukčného, kultúrneho aj sociálno-ekonomického. Takéto faktory totiž vplývajú na vývoj rodinného stavu obyvateľov. Priamo ho síce ovplyvňuje len sobášnosť, rozvodovosť a úmrtnosť, za týmito reprodukčnými faktormi je však mnoho iných vplyvov, ktoré pôsobia na rodinný stav sprostredkovane. Túto skutočnosť potvrdzujú aj vybrané faktory zo sčítania obyvateľov, ktoré boli predmetom predchádzajúceho hodnotenia. Pri všetkých sa preukázalo, že majú významný vplyv na rodinný stav obyvateľov.

Rodinný stav, podobne ako viaceré ďalšie populačné štruktúry, je na jednej strane výsledkom pôsobenia reprodukčných procesov, no zároveň je aj jedným z predpokladov demografického vývoja do budúcnosti. Treba ho brať do úvahy aj pri mnohých rozhodovacích a riadiacich procesoch v rôznych oblastiach spoločenského života.

Aj z predchádzajúceho hodnotenia je zrejmé, že rodinný stav obyvateľov má jednoznačný základný trend vyplývajúci z demografického vývoja. Je to znižovanie počtu a podielu osôb žijúcich v manželstve. Na úkor ženatých a vydatých osôb sa v populácii Slovenska už niekoľko desaťročí zvyšuje počet i podiel slobodných a rozvedených osôb. Je to dôsledok súčasného spôsobu života, keď kohabitácie čiastočne nahrádzajú manželstvá a manželské nezhody a krízy sa bežne riešia rozvodom. Z vývoja obyvateľov podľa rodinného stavu sa však dá usudzovať, že tento trend speje k svojmu naplneniu a intenzita vzniku kohabitácií aj rozvodovosť dosiahli svoj vrchol (pre rozvodovosť toto konštatovanie platí nesporne, pre kohabitácie existuje viac scenárov možného budúceho vývoja). Mohlo by to znamenať stabilizáciu štruktúry obyvateľov podľa rodinného stavu na súčasnej úrovni alebo dokonca mierne zvýšenie podielu ženatých a vydatých osôb v populácii. Takémuto trendu by mohol napomôcť aj pokles úmrtnosti a znižovanie nadúmrtnosti mužov, ktorý by spôsobil, že trvanie manželstiev by sa predĺžilo a podiel ovdovených osôb v populácii by sa znížil.

LITERATÚRA

[1] BLEHA, B. – GARAJOVÁ, A. – MÉSZÁROS, J. – PILINSKÁ, V. – ŠPROCHA, B. – VAŇO, B.: Populačný vývoj v krajoch a okresoch Slovenska od začiatku 21. storočia. Bratislava: INFOSTAT, 2019. 147 s. ISBN 978-80-89398-38-6.

[2] BLEHA, B. – PILINSKÁ, V. – ŠPROCHA, B. – VAŇO, B.: Populačný vývoj Slovenska v prierezo- a kohortnom pohľade. Bratislava: INFOSTAT, 2022. 150 s. ISBN 978-80-89398-47-8.

[3] SODB 2021 – Metodické vysvetlivky. Dostupné na: <https://www.scitanie.sk/vysvetlivky> [cit. 02-01-2023].

[4] VAŇO, B. (ed.): Populačný vývoj v Slovenskej republike 2011. Bratislava: INFOSTAT, 2012. 97 s. ISBN 978-80-89398-21-8.

[5] VAŇO, B. – HAVIAROVÁ, E.: Demografické trendy rómskej populácie. In: Vašečka, M. (ed.): Súhrnná správa o Rómoch na Slovensku. Bratislava: Inštitút pre verejné otázky, 2002, s. 475 – 502. ISBN 80-889-3541-5.

[6] Výsledky SĽDB 1991. Bratislava, Štatistický úrad SR. [cit. 02-01-2023]. Dostupné na: <https://www.scitanie.sk/o-scitani/udaje-z-minulych-scitani-od-1991>.

[7] Výsledky SODB 2001. Bratislava, Štatistický úrad SR. [cit. 02-01-2023]. Dostupné na: <https://www.scitanie.sk/o-scitani/udaje-z-minulych-scitani-od-1991>.

[8] Výsledky SODB 2011. Bratislava, Štatistický úrad SR. [cit. 02-01-2023]. Dostupné na: <https://www.scitanie.sk/o-scitani/udaje-z-minulych-scitani-od-1991>.

[9] Výsledky SODB 2021. [cit. 02-01-2023]. Dostupné na: <https://www.scitanie.sk/rozsirene-vysledky>.

RESUMÉ

Rodinný stav je dôležitou charakteristikou obyvateľstva. Poskytuje obraz o obyvateľstve z hľadiska reprodukčného, kultúrneho aj sociálno-ekonomického. Priamo ho ovplyvňuje sobášnosť, rozvodovosť a úmrtnosť, za týmito reprodukčnými faktormi je však mnoho iných vplyvov, ktoré pôsobia na rodinný stav sprostredkovane. Rodinný stav, podobne ako viaceré ďalšie populačné štruktúry, je na jednej strane výsledkom pôsobenia reprodukčných procesov, no zároveň je aj jedným z predpokladov demografického a spoločenského vývoja do budúcnosti.

Rodinný stav obyvateľov má v posledných desaťročiach jednoznačný základný trend vyplývajúci z demografického vývoja. Je to znižovanie počtu a podielu osôb žijúcich v manželstve. Na úkor ženatých a vydatých osôb sa v populácii Slovenska už niekoľko desaťročí zvyšuje počet aj podiel slobodných a rozvedených osôb. Podiel ovdovených zostáva stabilný a mení sa len minimálne.

Výrazné sú aj regionálne rozdiely v rodinnom stave obyvateľov, aj keď na krajskej úrovni sa neprejavujú v plnom rozsahu. Vo všetkých krajoch SR je zachovaná rovnaká základná štruktúra obyvateľov podľa rodinného stavu ako na celoslovenskej úrovni. Najviac obyvateľov je slobodných, nasledujú osoby žijúce v manželstve, ďalej rozvedení a najmenej je ovdovených. Jedinou výnimkou je Prešovský kraj, v ktorom je menej rozvedených ako ovdovených obyvateľov. Aj keď základné trendy sú na úrovni krajov rovnaké, regionálne rozdiely sa prejavujú v detailnejšom pohľade na rodinný stav obyvateľov.

Do štruktúrneho hodnotenia rodinného stavu obyvateľov sme zaradili ukazovatele, ktoré ovplyvňujú reprodukčné správanie obyvateľov a zároveň sa zisťujú v sčítaní obyvateľov, domov a bytov. Hodnotenie potvrdilo výrazné rozdiely v rodinnom stave obyvateľov podľa všetkých zvolených ukazovateľov.

Čo sa týka pohlavia, v populácii 15+ na Slovensku podiel slobodných mužov je vyšší ako podiel slobodných žien. U mužov je viac rozvedených ako ovdovených, u žien je situácia opačná, pričom podiel rozvedených aj ovdovených osôb je vyšší v populácii žien.

Vo veku do 34 rokov prevládajú v populácii Slovenska slobodné osoby. V strednom a staršom veku (35 – 79 rokov) je v populácii najviac ženatých, resp. vydatých osôb. Vo veku nad 80 rokov prevládajú medzi obyvateľmi ovdovené osoby. Najviac rozvedených v populácii Slovenska je vo veku okolo 50 rokov, tento rodinný stav však nie je v žiadnom veku dominantný.

Vysoký podiel slobodných je u obyvateľov bez detí a čiastočne aj s jedným dieťaťom. U obyvateľov s deťmi sú najpočetnejším rodinným stavom osoby žijúce v manželstve. Najviac rozvedených je u obyvateľov s jedným dieťaťom a najviac ovdovených v skupine obyvateľov s 5 a viac deťmi.

Obyvatelia so základným vzdelaním a bez vzdelania sú najčastejšie slobodní. U obyvateľov so stredoškolským a vysokoškolským vzdelaním prevláda rodinný stav ženatý, resp. vydatá. So zvyšujúcim sa vzdelaním klesá podiel ovdovených osôb v populácii a najviac rozvedených osôb je u obyvateľov so stredoškolským vzdelaním.

Obyvateľstvo slovenskej a maďarskej národnosti má podobnú štruktúru podľa rodinného stavu. U obyvateľov rómskej národnosti je veľmi vysoký podiel slobodných osôb, pri zvyšných hodnotených národnostiach je výraznejšia prevaha ženatých, resp. vydatých osôb.

Čo sa týka religiozity, medzi obyvateľmi, ktorí sa prihlásili ku štyrom najväčším cirkvám na Slovensku (rímskokatolícka, evanjelická a. v., gréckokatolícka a reformovaná) je najviac ženatých, resp. vydatých osôb, nasledujú slobodní, ovdovení a najmenej je rozvedených. V skupine ostatné cirkvi je viac rozvedených ako ovdovených. Výrazný rozdiel je v skupine obyvateľov bez vyznania. Podiel slobodných je vyšší ako podiel obyvateľov žijúcich v manželstve a podiel rozvedených je výrazne vyšší ako podiel ovdovených.

RESUME

Marital status is an important characteristic of population. It provides a picture of the population from the reproductive, cultural and socio-economic perspective. It is directly affected by marriage, divorce and mortality, but behind these reproductive factors there are many other influences affecting marital status indirectly. Marital status, like several other population structures, is on the one hand the result of reproductive processes, but at the same time it is also one of the prerequisites for demographic and social development in the future.

In recent decades, the marital status of the population has a clear basic trend resulting from the demographic development. It is a decrease in the number and the share of the married persons. At the expense of the married persons, the number and the share of single and divorced persons in the population of Slovakia has been increasing for several decades. The share of the widowed remains stable and changes only minimally.

The regional differences in the population's marital status are also significant, even if they are not fully manifested at or regional level. In all regions of the Slovak Republic, the same basic population structure according to family status is maintained as at the national level. The largest part of population is single, followed by the married persons, the divorced persons and the least numerous part of population are widowed persons. The only exception is the Prešov region, where there are fewer divorced persons than the widowed. Even if the basic trends are very similar at the regional level, the regional differences are manifested in a more detailed view of the marital status of population. In the structural assessment of the marital status, we included indicators affecting the reproductive behavior of the population and are also surveyed in the population census. The assessment confirmed significant differences in the marital status of population according to all selected indicators.

Regarding gender in the population aged 15 years and over the share of single men is higher than of single women. There are more divorced men than widowed in Slovakia. The reverse is true for women while the share of the divorced and the widowed persons is higher in the female population.

Single persons predominate in the population of Slovakia under the age of 34. In the middle and older age (35-79 years), the majority of the population is married. At the age of over 80, widowed persons predominate in population. Most divorced persons in the population of Slovakia are around age 50 years, but this marital status is not dominant at any age.

There is a high share of single persons among the population without any children and partly with one child. Among the population with children, the married people form the most numerous marital status. The largest number of divorced persons is among

persons with one child, and the majority of widowed persons can be found among persons with 5 or more children.

Population with basic education and without education is most often single. For population with secondary and higher education, the predominant marital status is married. With increasing education the share of widowed persons in the population decreases, and the highest number of divorced persons is among population with secondary education.

The Slovak and the Hungarians have a similar structure according to marital status. In the Roma population the share of single persons is very high. In the group of the remaining nationalities there is a significant predominance of married persons in the population.

As regards religiosity, among the population who have registered with the four largest churches in Slovakia (Roman Catholic, Evangelical, Greek Catholic and Reformed), the majority of population is married, followed by single and widowed persons and the smallest part are divorced persons. In the group of other religions, there are more divorced than widowed persons. There is a significant difference in the group of population without a religion. The share of unmarried persons is higher than the share of married people, and the share of divorced persons is significantly higher than the share of widowed persons.

PROFESIJNÝ ŽIVOTOPIS

Ing. Boris Vaňo vyštudoval Vysokú školu ekonomickú v Bratislave, následne absolvoval postgraduálne štúdium z demografie na Karlovej univerzite v Prahe. Od roku 1980 pracuje v Inštitúte informatiky a štatistiky ako výskumný pracovník v oblasti demografie. V rokoch 2000 – 2014 bol vedúcim Výskumného demografického centra. V období rokov 2006 – 2010 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

Milan TEREK
Vysoká škola manažmentu

CHARAKTERISTIKY VÝKONNOSTI PROCESU V METODOLÓGII SIX SIGMA

PROCESS PERFORMANCE CHARACTERISTICS IN SIX SIGMA METHODOLOGY

ABSTRAKT

V aplikáciách metodológie Six Sigma je pri hodnotení procesu nevyhnutné vhodne aplikovať charakteristiky výkonnosti procesu. Získané výsledky slúžia ako podklad pre rozhodovanie o projektoch Six Sigma. Cieľom príspevku je podať súhrnný prehľad týchto charakteristík s dôrazom na ich štatistický význam, prepojenia medzi nimi a spôsoby získavania dát na ich výpočet. Bude navrhnutý spôsob voľby rozsahu náhodného výberu a spôsob realizácie náhodného vyberania pri výpočte počtu nezhôd na milión príležitostí a budú uvedené interpretačné možnosti získaných výsledkov. Pochopenie štatistických súvislostí medzi charakteristikami výkonnosti procesu môže uľahčiť ich aplikácie v praxi. Aplikácia navrhnutého postupu pri náhodnom vyberaní môže umožniť presnejšie hodnotenia výkonnosti procesu.

ABSTRACT

In applying the Six Sigma methodology, it is essential to apply process performance characteristics when evaluating a process appropriately. The obtained results serve as a basis for decision-making on the Six Sigma projects. The aim of the paper is to provide a summary overview of these characteristics with an emphasis on their statistical significance, their interconnections, and the data collection methods for their calculation. The method of choosing the sample size and implementing a random sample when calculating the number of nonconformities per million opportunities will be proposed, and the interpretation possibilities of the obtained results will be presented. Understanding the statistical relationships among the process performance characteristics can facilitate their applications in practice. Applying the proposed random sampling procedure may facilitate more accurate evaluations of the process performance.

KĽÚČOVÉ SLOVÁ

Six Sigma, počet nezhôd na milión príležitostí, sigma úroveň kvality, tvorba náhodného výberu

KEYWORDS

Six Sigma, nonconformities per million opportunities, sigma quality level, creating a random sample

1. ÚVOD

Six Sigma je štruktúrovaný program alebo metodológia zlepšovania kvality vo všetkých aspektoch produktov firmy [8, s. xii]. V [15, s. 28] sa uvádza, že Six Sigma sa zameriava na znižovanie variability kľúčových charakteristík kvality produktu na úroveň, pri ktorej sú chyby alebo nezhody¹ extrémne nepravdepodobné. V [13, s. 1] je Six Sigma definovaná ako štruktúrovaný prístup, ktorý kombinuje chápanie potrieb zákazníka, racionálne využívanie štatistických metód zberu a analýzy dát a starostlivý

¹ Nezhodu možno definovať ako odchýlku od špecifikácie, normy alebo očakávania.

manažment procesov s cieľom zlepšovať podnikateľskú výkonnosť. Pojem Six Sigma sa prvýkrát objavil v polovici 80. rokov minulého storočia v technickom dokumente firmy Motorola s názvom *Six sigma mechanical design tolerance*. Hlavným cieľom Motoroly bolo znižovanie variability a počtu nezhôd. Za hlavný dôvod úspechu a popularity tejto metodológie v súčasnosti sa považuje využívanie rigorózneho metodológie na identifikáciu a elimináciu zdrojov variability [20, s. 9].

Prístup Six Sigma je projektovo orientovaný a zameraný na strategické ciele podnikania organizácie. Základnou filozofiou je zvyšovanie spokojnosti zákazníkov prostredníctvom eliminácie nezhôd a ich predchádzaniu a v dôsledku toho aj zvyšovanie ziskovosti podnikania. Hlavným účelom projektu Six Sigma je vyriešiť daný problém a tak prispieť k plneniu obchodných cieľov organizácie. Prístup využíva hlavne štatistické nástroje a odporúča sa jeho zosúladiť s plánmi manažmentu rizík a so všetkými inými aktivitami zameranými na prevenciu nezhôd.

S aktivitami Six Sigma je spojená procedúra DMAIC, ktorá zahŕňa päť fáz: definovať, merať, analyzovať, zlepšovať a kontrolovať. Vo fáze definovať treba identifikovať a definovať problém, na ktorom sa má pracovať. Vo fáze merať ide o meranie aktuálnej výkonnosti procesu, ktorý sa má zlepšiť. Vo fáze analyzovať sa majú zistiť hlavné príčiny nízkej výkonnosti. Fáza zlepšovať je zameraná na testovanie a štúdium potenciálnych riešení na vytvorenie robustného zlepšeného procesu. Napokon vo fáze kontrolovať sa zlepšený proces kontroluje prostredníctvom aplikácie štandardizovaného procesu, ktorý možno prevádzkovať a priebežne zdokonaľovať tak, aby bola zlepšená výkonnosť udržateľná v čase.

Hlas zákazníka (VOC – *voice of customer*) by mal poskytovať stálu spätnú väzbu počas celého trvania projektu Six Sigma. V kontexte projektu Six Sigma to môže byť sponzor projektu, interný alebo externý zákazník. Je dôležité, aby sa každý projekt Six Sigma začínal potrebami a očakávaniami zákazníkov. Následne by sa mali v každej fáze riešenia kontrolovať prebiehajúce aktivity projektu, aby sa potvrdilo, že sme sa neodchýlili od pôvodných očakávaní zákazníkov. Metodika Six Sigma by mala byť zameraná okrem spokojnosti zákazníkov aj na finančnú efektívnosť a bezpečnosť. Vo všetkých prípadoch by sa mal ako prvý krok vytvoriť účtovný model na vyhodnotenie finančnej prijateľnosti projektu. Výkonnosť skúmaného projektu by sa mala posudzovať z hľadiska efektívnosti pre zákazníka alebo na podnikanie.

Veličiny ktoré môžu ovplyvniť kvalitu finálneho produktu, sa súhrnne nazývajú charakteristiky kritické pre kvalitu (CTQCs – *critical-to-quality characteristics*). Ide o merateľné charakteristiky produktu alebo služby, ktorých výkonnostné normy alebo tolerančné hranice musia byť splnené, aby boli splnené požiadavky VOC. Podľa [9] predstavujú vlastnosti produktu alebo služby definované zákazníkom (interným alebo externým).

V článku si podrobnejšie všimneme fázu merať (*measure*) procedúry DMAIC, hlavne najdôležitejšie charakteristiky výkonnosti procesu. Vyjdeme z objasnenia štatistického významu Six Sigma a z modelu na meranie sigma úrovne kvality. Uvedieme spôsob výpočtu základnej charakteristiky výkonnosti procesu – počet nezhôd na milión príležitostí, všimneme si súvislosť medzi týmto ukazovateľom a sigma úrovňou kvality a opíšeme aj niektoré alternatívne možnosti určenia sigma úrovne kvality procesu na základe vypočítaného počtu nezhôd na milión príležitostí.

Pochopenie štatistických súvislostí medzi charakteristikami výkonnosti procesu môže uľahčiť ich aplikácie v praxi.

Pri výpočte výberového počtu nezhôd na milión príležitostí je nevyhnutné určiť rozsah výberu a spôsob vyberania. V [6], [19] sa napríklad uvádza len veľmi všeobecné pravidlo, že rozsah výberu by mal byť dostatočne malý, aby sa dal dobre zvládnuť, a zároveň dostatočne veľký, aby zachytil všetky špecifiká procesu. V [6] je uvedené: Uistite sa, že máte dostatočne veľký výber na to, aby sa mohli objaviť nezhody. Podrobnejšie návody na tvorbu výberu sa v literatúre neuvádzajú. V článku si podrobne všimneme problém spôsobu tvorby výberu, ktorý poskytne dáta na výpočet výberového počtu nezhôd na milión príležitostí, určenie jeho charakteru a rozsahu. Aplikácia navrhnutých postupov pri tvorbe výberu môže umožniť presnejšie a spoľahlivejšie hodnotenia výkonnosti procesu, čo môže viesť k prijímaniu lepších rozhodnutí o projektoch Six Sigma. Napokon uvedieme niektoré ďalšie charakteristiky výkonnosti procesu.

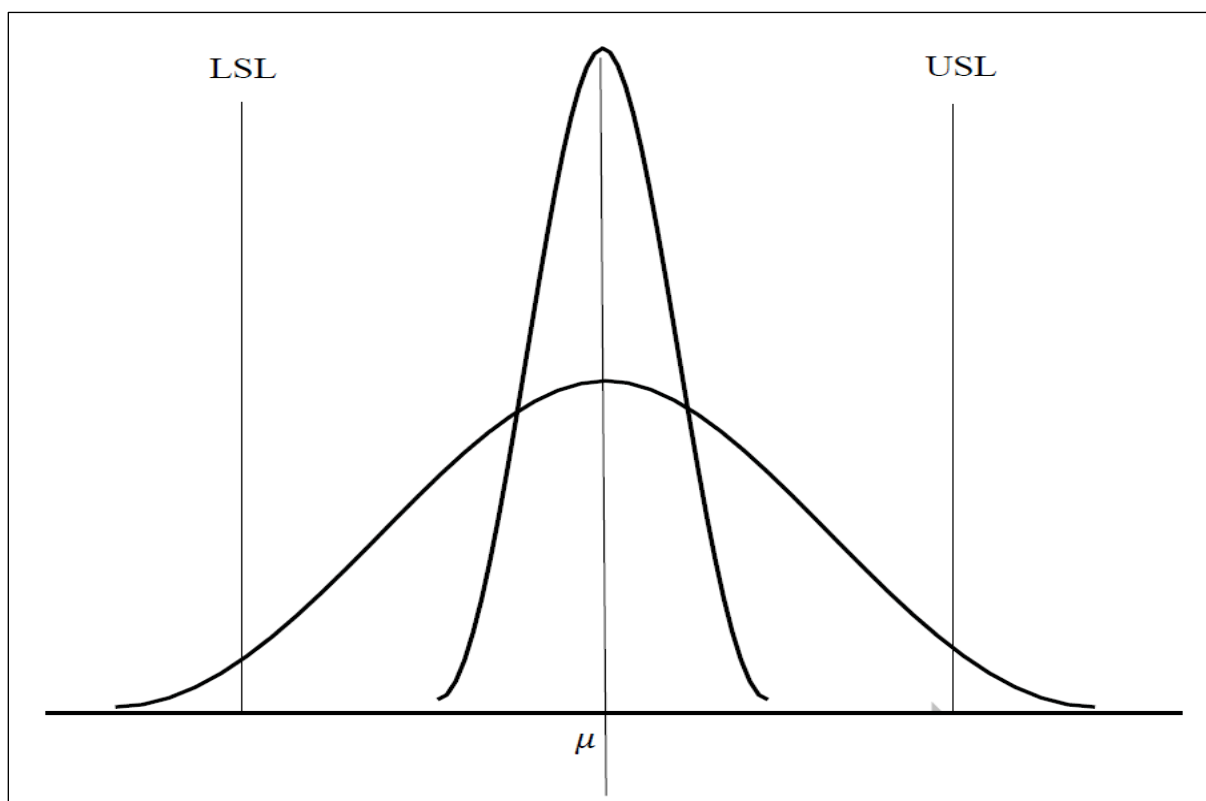
2. ŠTATISTICKÝ VÝZNAM SIX SIGMA A MODEL NA MERANIE SIGMA ÚROVNE KVALITY

Uvažujme najprv o jedinej charakteristike kritickej pre kvalitu produktu, resp. o jednoduchom procese, ktorý sa skladá z jediného kroku. Predpokladajme, že má normálne rozdelenie². Keď je jej hodnota vnútri tolerančných hraníc – LSL (*lower specification limit* – dolná tolerančná hranica) a USL (*upper specification limit* – horná tolerančná hranica), produkt je zhodný, keď nie je, produkt je nezgodný. Nech je cieľová hodnota charakteristiky kritickej pre kvalitu v strede tolerančného poľa. Na obrázku č. 1 sú grafy funkcie hustoty tejto charakteristiky, keď je centrovaná v strede tolerančného poľa. Prvý, menej špicatý graf je graf funkcie hustoty charakteristiky s normálnym rozdelením, so strednou hodnotou μ a so smerodajnou odchýlkou σ_1 , Druhý, špicatejší graf je graf funkcie hustoty tejto charakteristiky s normálnym rozdelením, so strednou hodnotou μ a so smerodajnou odchýlkou σ_2 , pričom $\sigma_2 < \sigma_1$. Plocha pod krivkami mimo tolerančných hraníc reprezentuje podiel nezgodných produktov. Je zrejmé, že keď sa variabilita charakteristiky zmenší zo σ_1 na σ_2 , zmenší sa aj podiel nezgodných produktov, ako to vidieť na obrázku č. 1.

Na obrázku č. 1 je stredná hodnota procesu μ v strede tolerančného poľa – rovná sa cieľovej hodnote. Pri návrhu koncepcie Six Sigma bol prijatý predpoklad, že stredná hodnota procesu je posunutá od cieľovej hodnoty o 1,5 smerodajnej odchýlky doprava alebo doľava. Je známe že výkonnosť procesu možno spoľahlivo predpovedať len vtedy, keď je proces stabilný, to znamená keď sa parametre rozdelenia pravdepodobnosti sledovanej charakteristiky v čase nemenia. V praxi sa však v procese môžu vyskytnúť poruchy, ktoré spôsobujú, že stredná hodnota procesu sa nerovná cieľovej hodnote. Podľa [2, s. 21], z dlhodobého hľadiska môže akumulácia malých posunov strednej hodnoty procesu viesť v najmenej priaznivom prípade k posunu strednej hodnoty procesu o 1,5 smerodajnej odchýlky doprava alebo doľava od cieľovej hodnoty. Podľa [10] je 1,5 sigma hodnotou odhadu posunu strednej hodnoty procesu medzi krátkym a dlhým obdobím.

² Metodológia Six Sigma využíva aproximatívne distribučné funkcie normálneho a Poissonovho rozdelenia, pričom ich využívanie je skôr matematickou nevyhnutnosťou ako prostou voľbou.

Obrázok č. 1: Grafy funkcie hustoty charakteristiky kritickej pre kvalitu pri jej odlišnej variabilite



Zdroj: vlastné spracovanie autora

Predpokladajme, že proces je regulovaný pomocou Shewhartových regulačných diagramov³. Je známe, že Shewhartove regulačné diagramy sú relatívne málo citlivé na malé posuny procesu, rádovo okolo 1,5 smerodajnej odchýlky alebo menej [15, s. 414]. To znamená, že v praxi sa často môžu vyskytovať situácie, že proces v nejakom kratšom období pracuje s posunom strednej hodnoty až o 1,5 smerodajnej odchýlky bez toho, že by sme dostali signál o narušení jeho stability. Proces teda v čase posunu strednej hodnoty o 1,5 smerodajnej odchýlky sa môže istý čas považovať za stabilný. Preto sa v koncepcii Six Sigma modeluje práve takéto správanie procesu. Ako všetky modely, ani tento určite nie je úplne presný, ale osvedčil sa ako vhodný pri úvahách o výkonnosti procesu⁴.

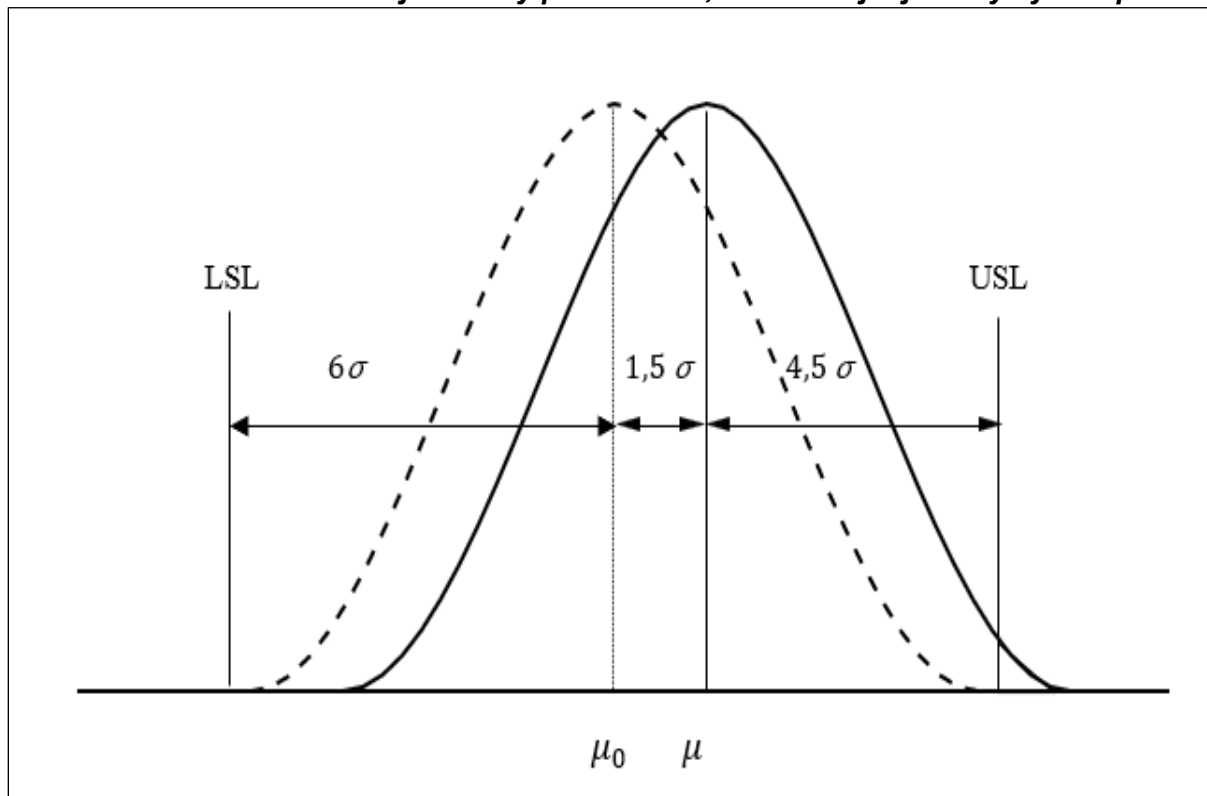
Na obrázku č. 2 je zobrazený graf funkcie hustoty procesu s normálnym rozdelením, so strednou hodnotou μ_0 a takou smerodajnou odchýlkou σ , že vzdialenosť medzi strednou hodnotou μ_0 a každou z tolerančných hraníc je 6σ spolu s posunom tohto grafu o $1,5\sigma$ doprava (stredná hodnota vzrástla z μ_0 na μ).

Všimnime si, ako možno počítať podiel zhodných jednotiek. Nech má charakteristika X normálne rozdelenie so strednou hodnotou $\mu = \mu_0 + 1,5\sigma$, kde μ_0 je cieľová hodnota charakteristiky v strede tolerančného poľa a so smerodajnou odchýlkou σ .

³ Podrobnejšie o Shewhartových regulačných diagramoch pozri v [15, 21].

⁴ Niektorí autori tento štandardný posun hodnotia ako „penalty value“, pretože neberie do úvahy skutočný stav zmeny strednej hodnoty v dlhom období.

Obrázok č. 2: Posun strednej hodnoty procesu o 1,5 smerodajnej odchýlky σ doprava



Zdroj: vlastné spracovanie

Všeobecne, horná a dolná tolerančná hranica USL a LSL sú:

$$\begin{aligned} \text{USL} &= \mu_0 + z\sigma \\ \text{LSL} &= \mu_0 - z\sigma \end{aligned}$$

kde z je hodnota náhodnej premennej Z s normovaným normálnym rozdelením a určuje počet smerodajných odchýlok σ , ktorý indikuje sigma úroveň kvality⁵. Napríklad pre $z = 3$, ide o tri sigma kvalitu, pre $z = 6$, o šesť sigma kvalitu a podobne. Pravdepodobnosť, že jednotka je zhodná je:

$$P(\text{LSL} \leq X \leq \text{USL})$$

Pri predpoklade, že $\mu = \mu_0 + 1,5\sigma$, budeme normovať USL a LSL:

$$z_{\text{USL}} = \frac{\mu_0 + z\sigma - (\mu_0 + 1,5\sigma)}{\sigma} = z - 1,5$$

$$z_{\text{LSL}} = \frac{\mu_0 - z\sigma - (\mu_0 + 1,5\sigma)}{\sigma} = -z - 1,5 = -(z + 1,5)$$

kde z_{USL} , z_{LSL} sú hodnoty náhodnej premennej Z s normovaným normálnym rozdelením.

⁵ Je známe, že každá hodnota z náhodnej premennej Z s normovaným normálnym rozdelením je vzdialenosť od strednej hodnoty meraná v smerodajných odchýlkach.

Zrejme:

$$\begin{aligned} P(LSL \leq X \leq USL) &= P(z_{LSL} \leq Z \leq z_{USL}) = P(-(z + 1,5) \leq Z \leq z - 1,5) = \\ &= P(Z \leq z - 1,5) - P(Z \leq -(z + 1,5)) = P(Z \leq z - 1,5) - P(Z \geq (z + 1,5)) = \\ &= P(Z \leq z - 1,5) - [1 - P(Z \leq (z + 1,5))] = \\ &= P(Z \leq z - 1,5) - 1 + P(Z \leq (z + 1,5)) = \Phi(z - 1,5) - 1 + \Phi(z + 1,5) \end{aligned}$$

kde Φ je distribučná funkcia normovaného normálneho rozdelenia. Jej hodnoty v rozličných bodoch budeme hľadať pomocou funkcie NORM.S.DIST v Exceli (postup pozri v [23, s. 39 – 40]).

Potom pravdepodobnosť, že jednotka je nezhodná je:

$$\begin{aligned} P(X \leq LSL) + P(X \geq USL) &= 1 - P(LSL \leq X \leq USL) = \\ &= 2 - \Phi(z - 1,5) - \Phi(z + 1,5) \end{aligned}$$

Rovnaký výsledok dostaneme pre $\mu = \mu_0 - 1,5\sigma$. Pravdepodobnosti, že jednotka je nezhodná budeme pre všetky $z > 0$ počítať podľa vzťahu:

$$2 - \Phi(z - 1,5) - \Phi(z + 1,5) \quad (1)$$

Všimnime si napríklad, ako možno vypočítať pravdepodobnosť, že jednotka je nezhodná pri tri sigma kvalite (pre $z = 3$), keď je stredná hodnota procesu posunutá od cieľovej hodnoty o 1,5 smerodajnej odchýlky doprava alebo doľava.

$$\begin{aligned} 2 - \Phi(z - 1,5) - \Phi(z + 1,5) &= 2 - \Phi(3 - 1,5) - \Phi(3 + 1,5) = 2 - \Phi(1,5) - \Phi(4,5) = \\ &= 2 - 0,933192799 - 0,999996602 = 0,066810599 \end{aligned}$$

Podiel nezhodných jednotiek je 6,6810599 %, čo je 66 810,599 nezhodných na milión jednotiek.

Pri štyri sigma kvalite je pravdepodobnosť, že jednotka je nezhodná:

$$2 - \Phi(2,5) - \Phi(5,5) = 2 - 0,993790335 - 0,999999981 = 0,006209684$$

Podiel nezhodných jednotiek je 0,6209684 %, čo je 6 209,684 nezhodných na milión jednotiek.

Pri päť sigma kvalite je pravdepodobnosť, že jednotka je nezhodná:

$$2 - \Phi(3,5) - \Phi(6,5) = 2 - 0,999767371 - 1 = 0,000232629$$

Podiel nezhodných jednotiek je 0,0232629 %, čo je 232,629 nezhodných na milión jednotiek.

Napokon pri šesť sigma kvalite máme pravdepodobnosť že jednotka je nezhodná:

$$2 - \Phi(4,5) - \Phi(7,5) = 2 - 0,999996602 - 1 = 0,000003398$$

Podiel nezhodných jednotiek je 0,0003398%, čo je 3,398 nezhodných na milión jednotiek.

Pravdepodobnosť že jednotka je zhodná možno vypočítať ako (1 – pravdepodobnosť že jednotka je nezhodná). Všetky výsledky sú v tab. č. 1.

Tabuľka č. 1: Sigma úrovne kvality

Sigma úroveň kvality	Pravdepodobnosť, že jednotka je nezhodná	Podiel nezhodných jednotiek (%)	Počet nezhodných na milión jednotiek	Pravdepodobnosť, že jednotka je zhodná
3 σ	0,066810599	6,6810599	66 810,599	0,933189401
4 σ	0,006209684	0,6209684	6 209,684	0,993790316
5 σ	0,000232629	0,0232629	232,629	0,999767371
6 σ	0,000003398	0,0003398	3,398	0,999996602

Zdroj: vlastné spracovanie

Tabuľku možno, samozrejme, podľa potreby rozšíriť⁶. Vráťme sa ešte raz k problému posunu strednej hodnoty o 1,5 smerodajnej odchýlky. Pri úrovni kvality 6 σ by podľa tohto modelu bolo približne 3,4 nezhodných na jeden milión jednotiek. Keď však uvažujeme o posune strednej hodnoty, proces nie je stabilný. Výkonnosť procesu možno spoľahlivo predpovedať len keď je proces stabilný – keď sa stredná hodnota a smerodajná odchýlka v čase nemenia⁷. Ak sa ale stredná hodnota v čase mení, nemožno hodnotu odhadu približne 3,4 nezhodných na jeden milión jednotiek považovať za veľmi spoľahlivú [15, s. 29]. Presnejší model však zatiaľ nie je k dispozícii.

3. CHARAKTERISTIKY VÝKONNOSTI PROCESU A ÚROVNE KVALITY

Uvedieme niektoré charakteristiky výkonnosti procesu. Stimulom na zlepšenie procesu sú najčastejšie tieto tri charakteristiky. Prvou je miera návratnosti RR (*return rate*), ktorá je definovaná ako počet vrátení alebo žiadostí o vrátenie produktu za určité obdobie, napr. mesiac, vydelený počtom expedovaných produktov. Ďalšou charakteristikou je počet hlásení problémov NPR (*number of problem reports*), ktorý je definovaný ako počet správ o problémoch s kvalitou produktu zaslaných zákazníkmi počas určitého obdobia, napríklad mesiaca. Bežne sa sleduje aj včasnosť dodania OTD (*on-time delivery*), ktorá meria včasnosť dodávok zákazníkovi. Definuje sa ako percento objednávok, ktoré sú doručené na miesta zákazníkov v súlade s ich požiadavkami. Nepretržité sledovanie týchto charakteristík naznačí, do akej miery ich treba zlepšiť [10]. Charakteristika náklady na nízku kvalitu COPQ (*cost of poor quality*) zachytáva náklady na odstraňovanie nezhôd, ktoré vznikli v dôsledku interných alebo externých príčin. V tomto príspevku si podrobnejšie všimneme charakteristiky výkonnosti procesu, ktorých výpočet je založený na dátach z náhodného výberu⁸.

3.1. POČET NEZHÔD NA MILIÓŇ PRÍLEŽITOSTÍ

Jeden finálny produkt (jednotka) sa bežne vyznačuje veľkým množstvom charakteristík kritických pre kvalitu alebo jeho produkcia prechádza viacerými krokmi procesu výroby. Sigma úroveň kvality pre rozličný počet charakteristík kritických

⁶ V tabuľke sa uvažuje o hornej aj dolnej tolerančnej hranici.

⁷ Myslí sa štatistická stabilita procesu.

⁸ Neuvažujeme o 100 % kontrole. Analýza možností využívania Six Sigma v tomto prípade by vyžadovalo samostatnú štúdiu.

pre kvalitu možno odhadovať pomocou ukazovateľa DPMO – počet nezhôd na milión príležitostí (*defects [nonconformities] per million opportunities*), ktorého hodnota vo výberovom súbore sa podľa [10] vypočíta takto:

$$DPMO = \frac{c}{n_{\text{units}} \cdot n_{CTQC}} \cdot 1\,000\,000 = DPO \cdot 1\,000\,000 \quad (2)$$

kde DPMO je výberový počet nezhôd na milión príležitostí⁹,
 c – celkový počet nezhôd vo výbere,
 n_{units} – rozsah výberu,
 n_{CTQC} – počet charakteristík kritických pre kvalitu,
 $n_{\text{units}} \cdot n_{CTQC}$ – celkový počet príležitostí,
 DPO – výberový priemerný počet nezhôd na jednu príležitosť.

Pre organizáciu ktorá chce aplikovať metodológiu Six Sigma, ide o jeden z najdôležitejších ukazovateľov.

Vráťme sa k pravdepodobnostnému modelu jednej charakteristiky kritickkej pre kvalitu – normálne rozdelenie so strednou hodnotou μ posunutou o 1,5 smerodajnej odchýlky doprava alebo doľava od cieľovej hodnoty μ_0 a so smerodajnou odchýlkou σ . Z tohto modelu sme vychádzali pri výpočte počtu nezhodných na milión jednotiek, ktoré sú vo štvrtom stĺpci tabuľky 1. Uvažujeme o jedinej charakteristike kritickkej pre kvalitu a podľa nej delíme jednotky na zhodné a nezahodné.

Uvažujme teraz namiesto jednotiek o príležitostiach na vznik nezhody. Každá príležitosť na vznik nezhody generuje alebo negeneruje jednu nezhodu. Uvažujme o rozdelení pravdepodobnosti nejakého hypotetického ukazovateľa kvality, ktorého hodnoty súhrnne zohľadňujú všetky uvažované charakteristiky kritické pre kvalitu. Nech je rozdelenie pravdepodobnosti tohto ukazovateľa rovnaké ako predtým – normálne so strednou hodnotou μ posunutou o 1,5 smerodajnej odchýlky σ doprava alebo doľava od cieľovej hodnoty μ_0 . Potom môžeme v tabuľke 1 zmeniť názvy stĺpcov od druhého po šiesty, takto: pravdepodobnosť, že príležitosť generuje nezhodu, podiel príležitostí, ktoré generujú nezhodu (%), počet príležitostí, ktoré generujú nezhodu na milión príležitostí, čo je rovnaké ako počet nezhôd na milión príležitostí (DPMO), Pravdepodobnosť, že príležitosť negeneruje nezhodu. Čísla vnútri tabuľky č. 1 zostanú nezmenené. Keď poznáme počet nezhôd na milión príležitostí, možno v takto modifikovanej tabuľke č. 1 nájsť príslušnú sigma úroveň kvality. Uvažovaný pravdepodobnostný model teda možno použiť na odhadovanie sigma úrovne kvality na základe vypočítanej hodnoty DPMO.

V modifikovanej tabuľke č. 1 si všimneme len prvý a štvrtý stĺpec a rozšírime ju tak, že budeme uvažovať o hodnotách $z = 2; 2,5; 3; 3,5; \dots, 6$. Výsledky sú v tabuľke č. 2¹⁰. V praxi sa najprv podľa vzťahu (2) vypočíta hodnota DPMO, potom sa v tabuľke 2 nájde zodpovedajúca približná sigma úroveň kvality.

Príklad č. 1. – Finálny produkt má 4 charakteristiky kritické pre kvalitu. Náhodne sa vybralo 100 finálnych produktov. Kontrolou sa na nich zistilo 7 nezhôd. Vypočítame

⁹ Myslia sa príležitosti na vznik nezhody.

¹⁰ Podrobnejšie tabuľky možno nájsť napríklad v [4] (Tabuľka S), [16] (Tabuľka T1).

DPMO podľa vzťahu (2):

$$DPMO = \frac{7}{100 \cdot 4} \cdot 1\,000\,000 = 0,0175 \cdot 1\,000\,000 = 17\,500$$

Vypočítanou hodnotou odhadneme hodnotu DPMO v základnom súbore. Keď výsledok porovnáme s hodnotami DPMO v tabuľke 2, vidno, že produkujeme medzi úrovňami 3,5 a 4 sigma. Hodnota 17 500 sa menej líši od hodnoty 22 750,419 ako od hodnoty 6 209,684. Môžeme prijať záver, že produkujeme približne na úrovni 3,5 sigma kvality.

Tabuľka č. 2: Sigma úroveň kvality (sigma skóre z) a DPMO

Sigma úroveň kvality (sigma skóre z)	Počet nezhôd na milión príležitostí DPMO
2	308 770,168
2,5	158 686,925
3	66 810,599
3,5	22 750,419
4	6 209,684
4,5	1349,899
5	232,629
5,5	31,671
6	3,398

Zdroj: vlastné spracovanie autora

V [10] v prílohe 1 na s. 28 – 29 je tabuľka A.1 — Sigma scores, v ktorej sú vypočítané hodnoty DPMO pre úrovne sigma kvality od 0 do 6, v členení po 0,01. Treba poznamenať, že hodnoty v tabuľke nie sú založené na výpočte pravdepodobnosti, že jednotka je nezhodná, podľa vzťahu (1). Vychádzajú z výpočtu pravdepodobnosti, že jednotka je nezhodná vzhľadom na hornú tolerančnú hranicu. Počítajú:

$$P(X \geq USL) = P(Z \geq z_{USL}) = P(Z \geq (z - 1,5)) = 1 - P(Z \leq (z - 1,5)) = 1 - \Phi(z - 1,5)$$

Výsledky v tabuľke A.1 sa len mierne líšia od výsledkov získaných na základe výpočtu pravdepodobnosti, že jednotka je nezhodná podľa vzťahu (1). Napríklad pre $z = 2$ je v tabuľke A.1 výsledok 308 538, v tabuľke č. 2 máme výsledok 308 770,168. Napríklad pre $z = 3$ je v tabuľke A.1 výsledok 66 807, v tabuľke č. 2 je 66 810,599 a podobne. Tieto malé rozdiely však prakticky neovplyvnia hodnotu odhadu sigma skóre. Pri hľadaní sigma skóre v príklade 1, možno v tabuľke A.1 nájsť hodnotu 17 429, ktorá sa najmenej líši od vypočítaného DPMO = 17 500. Hodnote 17 429 zodpovedá sigma skóre 3,61. Na základe tabuľky A.1 potom môžeme v príklade 1 prijať záver, že produkujeme približne na sigma úrovni 3,61.

Ak nie je k dispozícii podobná tabuľka, možno na určenie sigma skóre z použiť približný vzťah medzi sigma skóre z a DPMO [16, s. 23]:

$$z = 0,8406 + \sqrt{29,37 - 2,221 \cdot \ln DPMO} \quad (3)$$

3.1.1. PROCES VÝPOČTU DPMO

Uvedli sme, že v literatúre sme nenašli podrobnejšie informácie o tvorbe výberu, na základe ktorého sa vypočíta hodnota DPMO. Často sa neuvádza ani to, či má ísť o výsledok náhodného alebo zámerného vyberania. Niekedy je to len nepriamo naznačené. Napríklad v [26] sa uvádza, že získané výsledky sú hodnoty odhadov, čo indikuje nevyhnutnosť tvorby výberu náhodným vyberaním.

Pri praktickej realizácii vyberania treba ale nájsť odpovede na mnohé otázky. Napríklad: Čo je základný súbor? Je konečný alebo nekonečne veľký? Ako a kedy treba realizovať náhodné vyberanie? Aký by mal byť rozsah výberu? Postupne sa pokúsime na všetky uvedené otázky odpovedať. Vyjdeme z toho, že výberový počet nezhôd na milión príležitostí DPMO je hodnota odhadu, ktorou odhadujeme neznámy skutočný počet nezhôd na milión príležitostí v základnom súbore.

Je známe, že bodový odhad je výberová charakteristika, pomocou ktorej odhadujeme neznámu hodnotu nejakého parametra základného súboru. Výberové charakteristiky sú náhodné premenné, ktoré sú funkciou množiny náhodných premenných X_1, X_2, \dots, X_n – pozorovaní, ktoré tvoria náhodný výber, to znamená, že sú štatisticky nezávislé a rovnako rozdelené. Ďalej je známe, že pozorovania sú štatisticky nezávislé a rovnako rozdelené, keď sa realizuje náhodné vyberanie s opakovaním rozsahu n z konečného alebo nekonečne veľkého základného súboru alebo náhodné vyberanie bez opakovania z nekonečne veľkého základného súboru.

Základný súbor tvorí celá produkcia daného produktu. Je známe, že základný súbor, v ktorom je zaznamenanie každej jednotky nemožné alebo nerealizovateľné v reálnom čase sa považuje za nekonečne veľký, aj keď je v skutočnosti konečný¹¹. Za náhodný výber z nekonečne veľkého základného súboru sa považuje výber n jednotiek¹² zo základného súboru, ktorý sa získa tak, že sa rešpektujú dve podmienky: každá vybraná jednotka je z toho istého základného súboru¹³ a každá jednotka je vybraná nezávisle [1, s. 324]. Potom sú pozorovania štatisticky nezávislé a rovnako rozdelené náhodné premenné a možno používať bežné indukzívne štatistické metódy.

Spôsobilosť a výkonnosť procesu možno spoľahlivo predvídať, len ak je proces stabilný, to znamená, že parametre rozdelenia pravdepodobnosti procesu sa v čase nemenia. Keď je proces regulovaný, možno realizovať jeden náhodný výber rozsahu n , v čase, keď je proces stabilný. Aby sa zabezpečila platnosť podmienky, že všetky pozorovania sú z toho istého základného súboru, treba vybrať jednotky vyrobené v tom istom alebo približne tom istom čase¹⁴. Vzhľadom na to možno vybrať jednotky vyrobené v rade za sebou, teda v čo najkratšom čase. Vtedy možno predpokladať, že parametre rozdelenia pravdepodobnosti procesu sú rovnaké,¹⁵ a teda všetky pozorovania vo výbere sú z toho istého základného súboru. Rovnako sa postupuje pri odhadovaní krátkodobej spôsobilosti – spôsobilosti stroja [12, s. 280 – 281]. Spoločnosti zvyčajne hľadajú mieru zlepšenia procesu pomocou ukazovateľa DPMO, približne 10-krát každé dva roky [17, s. 22]. Podmienka nezávislosti by mala byť

¹¹ Pozri napríklad v [1], [22], [24], [25].

¹² Predpokladáme diskretný proces, v ktorom ide o výber n kusov produktu.

¹³ Myslí sa z toho istého rozdelenia pravdepodobnosti ukazovateľa kvality (procesu).

¹⁴ Použitý model totiž predpokladá možnosť malých posunov strednej hodnoty pri konštantnom rozptyle až do veľkosti posunu $1,5\sigma$, teda proces typu C (podrobnejšie pozri v [12, s. 47 – 49].

¹⁵ Respektíve približne rovnaké.

splnená tak, že jednotky sa vyrábajú nezávisle, a teda výrobu každej jednotky možno považovať za realizáciu nezávislého náhodného pokusu.

Situácia je zložitejšia, ak chceme odhadovať počet nezhôd na milión príležitostí v neregulovanom procese. Keď proces nie je regulovaný, nevieme jednoznačne posúdiť jeho stabilitu. V [15, s. 374] sa uvádza, že ak proces nie je stabilný, štatistické vlastnosti indexov výkonnosti procesu sa nedajú určiť, a tak nie je možné robiť žiadne platné závery o ich skutočných hodnotách v základnom súbore. To samozrejme platí aj o DPMO. Ak bola jeho hodnota vypočítaná z dát z náhodného výberu z nestabilného procesu¹⁶, nemožno na základe nej urobiť platný záver o hodnote DPMO v základnom súbore. Ak nemáme žiadne informácie o stabilite procesu, môže pri odhadovaní výkonnosti procesu ísť len o prieskumnú analýzu. Hodnotou DPMO získanou na základe dát z výberu nemožno odhadovať neznámy skutočný počet nezhôd na milión príležitostí v základnom súbore. Ide len o hodnotu opisnej charakteristiky výkonnosti procesu vo výberovom súbore. Ak robíme na základe vypočítanej hodnoty DPMO nejaké závery o skutočnom počte nezhôd na milión príležitostí v základnom súbore, ide len o subjektívne ohodnotenie.

Prípadne môžeme mať nejaké indície o stabilite procesu. Predstavte si situáciu, že proces nie je regulovaný, ale za dlhšie obdobie sa napríklad miera návratnosti RR v jednotlivých mesiacoch, počet hlásení problémov NPR v jednotlivých mesiacoch, prípadne náklady na nízku kvalitu COPQ v jednotlivých mesiacoch prakticky nemenia. To by mohlo indikovať stabilitu procesu výroby na nejakej sigma úrovni kvality. Aj v tomto prípade však treba byť pri interpretácii charakteristík výkonnosti procesu veľmi opatrný.

Pri určovaní rozsahu náhodného výberu sa treba oprieť o znalosti o charaktere procesu. V [11, s. 2] sa uvádza, že metódy na meranie výkonnosti strojov uvedené v dokumente, možno použiť aj pri vykonávaní auditov procesu, pričom rozsah výberu by v tomto prípade mal byť aspoň 50. Podobne možno uvažovať pri odhadovaní počtu nezhôd na milión príležitostí.

Existuje aj alternatívna možnosť určenia rozsahu výberu. Ak chceme počítat a interpretovať intervaly spoľahlivosti pre DPMO, možno postupovať takto. Uvažujeme o príležitostiach na generovanie nezhody a predpokladáme že jedna príležitosť môže generovať práve jednu nezhodu. Potom možno vypočítané DPO chápať ako hodnotu výberového podielu príležitostí ktoré generujú nezhodu, z celkového počtu príležitostí vo výbere. V [25] na s. 5 – 7 sú uvedené vzťahy na výpočet intervalu spoľahlivosti pre podiel a výpočet potrebného rozsahu výberu. Podľa nich možno vypočítat potrebný rozsah výberu a interval spoľahlivosti pre DPO v základnom súbore. Po vynásobení vypočítaného intervalu spoľahlivosti jedným miliónom, dostaneme príslušný interval spoľahlivosti pre DPMO. Predpokladajme, že v príklade 1 sme realizovali predbežný výber napríklad rozsahu 30 jednotiek a hodnota výberového podielu DPO vyšla 0,025. Touto hodnotou odhadneme podiel π vo vzťahu (3) v [25]. Ak napríklad stanovíme najväčšiu chybu $d = 0,02$, a spoľahlivosť $(1 - \alpha) = 0,95$, vyjde potrebný počet príležitostí $n = 234,1 \approx 235$. Z toho potrebný rozsah výberu¹⁷ je $235/4 = 58,75 \approx 59$. Po realizácii doplnkového výberu rozsahu $(59 - 30) = 29$ sa na základe

¹⁶ Nie je zaručené, že pozorovania sú z toho istého základného súboru.

¹⁷ Uvažujeme o štyroch charakteristikách kritických pre kvalitu.

celého výberu rozsahu 59 jednotiek vypočíta nová hodnota DPO. Povedzme, že by vyšlo $DPO = 0,021$. Podľa vzťahu (1) v [25] možno vypočítať 95 % interval spoľahlivosti pre DPO ¹⁸. Ten vyjde (0,00267; 0,03933). Po jeho vynásobení jedným miliónom dostaneme 95 % interval spoľahlivosti pre $DPMO$. Ten vyjde (2 670; 39 330). V tabuľke A.1 v [10] nájdeme príslušné sigma úrovně kvality. So spoľahlivosťou 0,95 produkujeme na sigma úrovni kvality z intervalu (3,26; 4,29).

V [6] sa uvádza, že výber má byť dostatočne veľký na to, aby sa mohla objaviť nezhoda. Ďalšou podmienkou na určenie rozsahu výberu by teda malo byť splnenie požiadavky, aby sa na jednotkách náhodného výberu vyskytla aspoň jedna nezhoda. Všeobecne, pri implementácii náhodného vyberania z nekonečne veľkého základného súboru treba postupovať veľmi uvážlivo. Každý konkrétny prípad môže vyžadovať rozličnú procedúru vyberania [1, s. 324].

Pri určovaní charakteristík kritických pre kvalitu sa väčšinou postupuje tak, že sa vytvorí zoznam potenciálnych nezhôd, ktoré môžu zákazníkom prekážať. Pritom sa treba zamerať na nezhody, ktoré sa môžu bežne vyskytovať, nie na zriedkavé nezhody. Nakoniec sa odporúča zlúčenie podobných nezhôd do jednej kategórie. Pri tvorbe zoznamu treba stále pamätať na hlas zákazníka. Niektoré vlastnosti produktov môžu byť pre zákazníkov dôležité, iné môžu byť príjemné. Treba sa zamerať na charakteristiky, ktoré ovplyvnia spokojnosť zákazníkov a sú merateľné. Ak do zoznamu pridáte aj charakteristiky, ktoré pre zákazníka nie sú dôležité, výsledkom bude nižšia hodnota $DPMO$, vyššia sigma úroveň kvality a falošný dojem, že výkonnosť procesu je vyššia, než je v skutočnosti [6]. Všeobecne ale neexistuje jasný návod, ako určiť počet charakteristík kritických pre kvalitu [7]. Nakoniec sa vypočíta hodnota $DPMO$ podľa vzťahu (2) alebo interval spoľahlivosti pre $DPMO$. Ak sa výber realizoval v čase, keď bol proces stabilný, odhadujeme vypočítanou hodnotou $DPMO$, resp. intervalom spoľahlivosti pre $DPMO$, neznámy skutočný počet nezhôd na milión príležitostí v základnom súbore.

Príklad 2. – Pri výrobe súčiastky elektrického zariadenia rozoznávame zhodné a nezgodné súčiastky. Uvažujeme teda o jedinej charakteristike kritickej pre kvalitu. Chceme odhadnúť sigma úroveň kvality procesu. Proces výroby je regulovaný pomocou Shewhartových regulačných diagramov. V priebehu realizácie náhodného vyberania rozsahu $n_{\text{units}} = 100$ regulačné diagramy nesignalizovali narušenie stability procesu¹⁹. V náhodnom výbere sa našli 2 nezgodné súčiastky. Vypočítame hodnotu výberového $DPMO$, ktorou bodovo odhadneme hodnotu $DPMO$ v základnom súbore:

$$DPMO = \frac{2}{100 \cdot 1} \cdot 1\,000\,000 = 20\,000$$

Podľa tabuľky č. 2 je sigma skóre medzi 3,5 a 4. Podľa tabuľky A.1 v [10] je sigma skóre približne 3,55.

Vráťme sa na chvíľu k požiadavke, aby sa na jednotkách náhodného výberu našla aspoň jedna nezhoda. Všimnime si na dátach z príkladu 2, ako by sa menila sigma úroveň kvality, keby sme 2 nezgodné súčiastky našli vo výbere rozsahu 1 000

¹⁸ Za n sa dosadí počet príležitostí.

¹⁹ Ide o samostatný náhodný výber na posúdenie sigma úrovně kvality, nie o náhodné výbery realizované v rámci štatistickej regulácie procesov.

súčiastok. Príslušná sigma úroveň kvality by bola 4,38. Výber je 10-krát väčší, sigma úroveň kvality vzrástla z 3,55 na 4,38. Keby sme našli 1 nezhodnú súčiastku vo výbere rozsahu 100 a 1000 súčiastok, príslušné sigma úrovne kvality by boli 3,83 a 4,59. Je pritom jedno, či nenájdeme žiadnu nezhodnú súčiastku vo výbere rozsahu 100, 1 000 alebo ľubovoľnom inom. Model je pre $c = 0$ nepoužiteľný. Preto sa kvôli zachovaniu rozlišovacej schopnosti kritéria zdá užitočné požadovať, aby bola vo výbere nájdená aspoň jedna nezhoda²⁰. Ak by sme v príklade 2 vo výbere 100 jednotiek nenašli žiadnu nezhodnú jednotku, pokračovali by sme vo výbere, pokiaľ by sa nezhodná jednotka neobjavila, až po nejakú určenú hranicu, napríklad 500 jednotiek. Je možné, že ani posledná jednotka nebude nezhodná a teda celkový výsledok bude 0 nezhôd vo výbere. V [26] odporúčajú v takomto prípade použiť nejaké informácie o frekvencii výskytu nezhôd z minulosti. Prípadne možno celé náhodné vyberanie zopakovať.

Všeobecne možno hodnotu výberového DPMO ľubovoľne znížiť pridaním ďalších charakteristík kritických pre kvalitu, preto by sa s odhadmi DPMO pre viac ako jednu charakteristiku malo zaobchádzať veľmi opatrne [18, s. 171]. Nazdávame sa, že ak sa pomocou DPMO pravidelne hodnotí nejaký proces v jednej firme, počet charakteristík kritických pre kvalitu sa nemení a všetky uvedené predpoklady na kvalitné odhadovanie sú dodržané, nemožno mať proti tejto charakteristike vážne výhrady.

3.2. POČET NEZHÔD NA JEDNOTKU

Často môže byť zaujímavý počet nezhôd na jednotku v základnom súbore, ktorý odhadujeme z dát náhodného výberu hodnotou výberového priemerného počtu nezhôd na jednotku DPU:

$$DPU = \frac{c}{n_{unit}} \quad (4)$$

Keď dosadíme DPU zo vzťahu (4) za $\frac{c}{n_{unit}}$ do vzťahu (2), dostaneme:

$$DPMO = \frac{DPU}{n_{CTQC}} \cdot 1\,000\,000$$

Z porovnania ostatného vzťahu a vzťahu (2) je zřejmé, že:

$$DPO = \frac{DPU}{n_{CTQC}} \quad (5)$$

a

$$DPU = DPO \cdot n_{CTQC}$$

Niekedy nás zaujímajú pravdepodobnosti, že na jednotke sa vyskytne nejaký konkrétny počet nezhôd. Pri hľadaní takýchto pravdepodobností možno využiť Poissonovo rozdelenie. Všeobecne, Poissonovo rozdelenie môže slúžiť ako model počtu úspechov, ktoré sa vyskytnú počas daného časového intervalu alebo v špecifikovanej oblasti. Predpokladá sa [14, s. 163]:

1) počty úspechov, ktoré sa vyskytnú v neprekrývajúcich sa časových intervaloch alebo oblastiach, sú nezávislé,

²⁰ Tým sa, samozrejme, náhodné vyberanie nemení na zámerné vyberanie.

2) pravdepodobnosť výskytu úspechu vo veľmi krátkom časovom intervale alebo vo veľmi malej oblasti je úmerná dĺžke časového intervalu alebo veľkosti oblasti,

3) pravdepodobnosť, že v krátkom časovom intervale alebo v malej oblasti sa vyskytne viac ako jeden úspech je zanedbateľne malá.

Pravdepodobnostná funkcia Poissonovho rozdelenia je:

$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{pre } k = 0, 1, 2, \dots$$

kde $\lambda > 0$ je stredná hodnota počtu úspechov ktoré sa vyskytnú v danom časovom intervale alebo v danej oblasti,

k – počet úspechov ktoré sa vyskytnú v danom časovom intervale alebo v danej oblasti.

Keď za úspech považujeme výskyt nezhody, môžeme počet nezhôd, ktoré sa vyskytnú na jednej jednotke produktu modelovať Poissonovým rozdelením. Strednú hodnotu λ odhadneme hodnotou výberového priemerného počtu nezhôd na jednotku – DPU. Nech $k = 0, 1, 2, \dots$ je počet nezhôd, ktoré sa vyskytnú na jednej jednotke. Potom pravdepodobnosť že na jednej jednotke sa vyskytne k nezhôd je [2, s. 81], [3, s. 132]:

$$p_k = P(X = k) = \frac{\text{DPU}^k}{k!} e^{-\text{DPU}}$$

Pravdepodobnosť, že na jednotke sa nevyskytne žiadna nezhoda (obsahuje $k = 0$ nezhôd), je:

$$p_0 = P(X = 0) = \frac{\text{DPU}^0}{0!} e^{-\text{DPU}} = e^{-\text{DPU}} \quad (6)$$

a pravdepodobnosť že na jednotke sa vyskytne aspoň jedna nezhoda, je:

$$1 - p_0 = 1 - e^{-\text{DPU}} \quad (7)$$

Príklad 1 – pokračovanie 1. – Predpokladajme, že keď sa na jednotke nevyskytne žiadna nezhoda, jednotka je zhodná, keď sa na nej vyskytne aspoň jedna nezhoda, jednotka je nezhodná. V príklade vypočítame pravdepodobnosť, že jednotka je zhodná a pravdepodobnosť, že je nezhodná.

Najprv vypočítame podľa vzťahu (4) priemerný počet nezhôd na jednotku DPU:

$$\text{DPU} = \frac{c}{n_{\text{unit}}} = \frac{7}{100} = 0,07$$

alebo

$$\text{DPU} = \text{DPO} \cdot n_{\text{CTQC}} = 0,0175 \cdot 4 = 0,07$$

Pravdepodobnosť že jednotka je zhodná vypočítame podľa vzťahu (6):

$$p_0 = P(X = 0) = e^{-0,07} = 0,932393819$$

Pravdepodobnosť že jednotka je nezhodná vypočítame podľa vzťahu (7):

$$1 - p_0 = 0,067606181$$

3.3. CELKOVÁ VÝŤAŽNOSŤ RTY

Výťažnosť²¹ FPY_i (*first-pass yield*) je pravdepodobnosť, že jednotka prejde i -tým krokom procesu výroby bez nezhôd. Celková výťažnosť RTY (*rolled throughput yield*) je pravdepodobnosť, že jednotka prejde všetkými krokmi procesu bez nezhôd [2, s. 82]. Celková výťažnosť RTY sumarizuje dáta o počte nezhôd na príležitosť pre proces alebo produkt [17, s. 484].

Keď uvažujeme o jednom kroku procesu a o jednom druhu nezhody, jednotka neprejde i -tým krokom procesu, počet charakteristík kritických pre kvalitu je $n_{CTQC} = 1$ a $DPO_i = DPU_i$. Hodnotou DPU_i možno odhadovať pravdepodobnosť, že jednotka je v i -tom kroku nezhodná. Výťažnosť FPY_i potom vypočítame takto:

$$FPY_i = 1 - DPU_i \quad (8)$$

Možno postupovať aj inak. Priemerným počtom nezhôd na jednotku v i -tom kroku procesu DPU_i odhadneme strednú hodnotu Poissonovho rozdelenia. Potom možno výťažnosť i -tého kroku procesu FPY_i vypočítať ako pravdepodobnosť že na jednotke sa nevyskytne žiadna nezhoda [8, s. 239]:

$$FPY_i = e^{-DPU_i} \quad (9)$$

Keď naopak poznáme FPY_i , možno vypočítať DPU_i . Vzťah na jeho výpočet možno odvodiť takto:

$$\ln FPY_i = -DPU_i \ln e$$

a z toho:

$$DPU_i = -\ln FPY_i$$

Ak má proces výroby produktu n krokov. Celková výťažnosť RTY sa vypočíta takto (predpokladá sa, že kroky procesu sú nezávislé²²):

$$RTY = \prod_{i=1}^n FPY_i \quad (10)$$

Hodnotou $DPU = (1 - RTY)$ možno odhadnúť pravdepodobnosť, že jednotka je nezhodná.

Príklad 3. – Proces výroby produktu sa skladá z troch nezávislých krokov. Z 1 000 náhodne vybraných produktov bolo v prvom kroku vyradených²³ 10 jednotiek,

²¹ Podľa [5].

²² Ak sa vyskytnú skupiny závislých krokov, možno kroky v každej skupine agregovať (ak to je možné) do jedného kroku.

²³ Vyradená jednotka mohla ísť do odpadu, na opravu alebo prepracovanie.

v druhom kroku 20 jednotiek a v treťom kroku sa vyradilo 19 jednotiek. Vypočítame DPU_i , FPY_i , RTY , DPU , DPO a $DPMO$. Alternatívne vypočítame FPY_i aj podľa vzťahu (9) a následne RTY , DPU , DPO a $DPMO$.

Počty nezhôd na jednotku jednotlivých krokov procesu odhadujeme hodnotami:

$$DPU_1 = \frac{10}{1000} = 0,0100; \quad DPU_2 = \frac{20}{990} = 0,0202; \quad DPU_3 = \frac{19}{970} = 0,0196$$

Výťažnosti v jednotlivých krokoch procesu odhadujeme hodnotami:

$$FPY_1 = 1 - 0,0100 = 0,9900; \quad FPY_2 = 1 - 0,0202 = 0,9798; \\ FPY_3 = 1 - 0,0196 = 0,9804$$

Odhadnutú celkovú výťažnosť RTY vypočítame podľa vzťahu (10):

$$RTY = \prod_{i=1}^3 FPY_i = 0,9900 \cdot 0,9798 \cdot 0,9804 = 0,9510$$

a

$$DPU = 1 - RTY = 0,049$$

Odhadnutá pravdepodobnosť, že jednotka je zhodná, je 0,9510 a odhadnutá pravdepodobnosť že jednotka je nezhodná je 0,049.

Keďže $n_{CTQC} = 1$, máme:

$$DPO = DPU = 0,049$$

Vypočítame $DPMO$:

$$DPMO = DPO \cdot 1\,000\,000 = 49\,000$$

Rovnakú hodnotu $DPMO$ dostaneme, keď do vzťahu (2) dosadíme $c = 10 + 20 + 19 = 49$, $n_{units} = 1000$ a $n_{CTQC} = 1$. Tejto hodnote $DPMO$ zodpovedá podľa tabuľky A.1 v [10], sigma úroveň kvality 3,15.

Keby sme FPY_i počítali podľa vzťahu (9) a potom RTY podľa vzťahu (10), vyšlo by $RTY = 0,9514$. Z toho potom $DPU = 1 - 0,9514 = 0,0486$, $DPMO = 48\,600$ a sigma úroveň kvality vyjde 3,16. Všetky výsledky sa líšia len minimálne. Je nakoniec jedno ktorý postup zvolíme. Dôležité je, aby sme pri opakovanom posudzovaní jedného procesu zachovali vždy rovnaký postup.

Niekedy môže byť užitočný výpočet normalizovanej výťažnosti (*normalized yield*) na základe RTY . Pre n -krokový proces sa vypočíta takto:

$$FPY_N = \sqrt[n]{RTY}$$

kde FPY_N je priemerná výťažnosť (geometrický priemer) v jednom kroku procesu.

Ak výťažnosti v jednotlivých krokoch procesu nie sú rovnaké, potom FPY_N je požadovaná výťažnosť najhoršieho z nich. Napríklad pre 10-krokový proces s požadovaným $RTY = 0,9999$ je najhoršia prijateľná výťažnosť v každom kroku procesu $\sqrt[10]{0,9999} = 0,99999$.

Žiaľ, nájdenie RTY často nie je také jednoduché, ako je opísané vyššie. V praxi sa len zriedka nájde postupnosť procesných krokov, ktoré do seba presne zapadajú. Obyčajne totiž máme rôzne toky dodávateľov, s rozličnými objemami a výnosmi. Niektoré kroky sa vykonajú vždy, niektoré nie. Testovacie a kontrolné body nemusia vždy poskytovať úplne presné výsledky. Niekedy treba vykonať opravy produktov počas procesu ich výroby. Niekedy možno sledovať konkrétnu dávku vstupov v procese a výsledky procesu po každom kroku, ale niekedy to je príliš zložité. Výrobné a informačné systémy nie sú často navrhnuté tak, aby poskytovali presné výsledky. Alternatívu ponúkajú simulačné modely a príslušný softvér. Pomocou simulácie a príslušného softvéru možno jednotlivé procesné kroky modelovať a spájať. Na simulačnom modeli možno potom monitorovať výsledky procesu tak často, ako je to potrebné [17, s. 488].

4. ZÁVER

Cieľom príspevku bolo podanie súhrnného prehľadu charakteristík na meranie výkonnosti procesu v metodológii Six Sigma s dôrazom na ich štatistický význam a vzájomné prepojenia a podrobnejšia analýza problému tvorby výberu, ktorý poskytne dáta na výpočet počtu nezhôd na milión príležitostí a súvisiacich ukazovateľov.

Vyšli sme z objasnenia štatistického významu Six Sigma a z modelu na meranie sigma úrovně kvality. Uviedli sme spôsob výpočtu základnej charakteristiky výkonnosti procesu *počet nezhôd na milión príležitostí*. Bola opísaná súvislosť medzi týmto ukazovateľom a sigma úrovňou kvality a uvedené niektoré alternatívne možnosti určenia sigma úrovně kvality procesu na základe vypočítaného počtu nezhôd na milión príležitostí. V časti proces výpočtu DPMO bol analyzovaný spôsob tvorby výberu a možnosti určenia jeho rozsahu. Keď je proces regulovaný, možno realizovať jeden náhodný výber rozsahu n v čase, keď je proces stabilný. Aby sa zabezpečila platnosť podmienky, že všetky jednotky sú z toho istého základného súboru, treba vybrať jednotky vyrobené v tom istom alebo približne v tom istom čase. Vzhľadom na to možno vybrať jednotky vyrobené v rade za sebou, teda v čo najkratšom čase. Rozsah výberu by mal byť aspoň 50 jednotiek, v prípade odhadovania DPMO pomocou intervalu spoľahlivosti, možno rozsah výberu pre určenú spoľahlivosť a najväčšiu chybu, vypočítať. Ďalšou podmienkou by malo byť splnenie požiadavky, aby sa na jednotkách náhodného výberu vyskytla aspoň jedna nezhoda. Všeobecne, pri implementácii náhodného vyberania z nekonečne veľkého základného súboru treba postupovať veľmi uvažlivo. Každý konkrétny prípad môže vyžadovať rozličnú procedúru vyberania.

Keď nemáme žiadne informácie o stabilite procesu, môže pri predvídaní výkonnosti procesu ísť len o prieskumnú analýzu. Pri tvorbe náhodného výberu možno postupovať ako v prípade stabilného procesu, ale hodnotou DPMO získanou z výberu nemožno odhadovať neznámy skutočný počet nezhôd na milión príležitostí v základnom súbore. Ide len o hodnotu opisnej charakteristiky výkonnosti procesu vo výberovom súbore. Ak takto získanou hodnotou DPMO napriek tomu odhadneme

počet nezhôd na milión príležitostí v základnom súbore, ide len o subjektívne ohodnotenie.

Ak máme o stabilite procesu len nejaké indície, napríklad, že v neregulovanom procese sa za dlhšie obdobie prakticky nemení miera návratnosti RR v jednotlivých mesiacoch, počet hlásení problémov NPR v jednotlivých mesiacoch, prípadne náklady na nízku kvalitu COPQ, môže to indikovať stabilitu procesu výroby na nejakej sigma úrovni kvality. Aj v tomto prípade však treba byť pri interpretácii charakteristík výkonnosti procesu veľmi opatrný. Zaujímavé informácie o výkonnosti procesu môžu poskytnúť aj ukazovatele DPU a RTY.

Pochopenie štatistických súvislostí medzi charakteristikami výkonnosti procesu môže značne uľahčiť proces ich aplikácie v praxi. Uplatnenie navrhnutých postupov pri tvorbe výberu môže umožniť presnejšie a spoľahlivejšie hodnotenia výkonnosti procesu. To všetko môže viesť k prijímaniu lepších rozhodnutí o projektoch Six Sigma.

LITERATÚRA

- [1] ANDERSON, D. R. – SWEENEY, D. J. – WILLIAMS, T. A. – CAMM, J. D. – COCHRAN, J. J. – FRY, M. J. – OHLMANN, J. W.: Statistics for Business and Economics. 14e Edition. Boston: Cengage Learning, Inc., 2020. 1119 s. ISBN 978-1-337-90106-2.
- [2] BASS, I.: Six Sigma Statistics with Excel and Minitab. New York: McGraw-Hill,, 2007. 374 s. ISBN 0-07-154268-X.
- [3] BASS, I. – LAWTON, B.: Lean Six Sigma Using SigmaXL and Minitab. New York: McGraw-Hill, 2009. 362 s. ISBN 978-0-07-162621-7.
- [4] BREYFOGLE III, F. W.: Implementing Six Sigma. Smarter Solutions Using Statistical Methods. New York: J. Wiley and Sons, 2003. 790 s. ISBN 978-0-471-26572-6.
- [5] BURIETA, J.: Kvalita v administratívnych procesoch [cit. 2023-05-16]. Dostupné na: <https://www.ipaslovakia.sk/clanok/kvalita-v-administrativnych-procesoch>.
- [6] Defects Per Million Opportunities (DPMO) [cit. 2023-05-16]. Dostupné na: <https://www.isixsigma.com/dictionary/defects-per-million-opportunities-DPMO/>.
- [7] DPMO (Defects per Million Opportunities): Simplified, With Example [cit. 2023-05-16]. Dostupné na: <https://www.knowledgehut.com/blog/quality/DPMO>.
- [8] EL-HAIK, B. – AL-AOMAR, R.: Simulation-Based Lean Six-Sigma and Design for Six-Sigma. Hoboken: J. Wiley and Sons, 2006. 404 s. ISBN 13: 978-0-471-69490-8.
- [9] Formulate Critical-to-Quality Characteristics [cit. 2023-05-16]. Dostupné na: <https://www.informit.com/articles/article.aspx?p=412351&seqNum=4>.
- [10] ISO 13053 - 1 Quantitative methods in process improvement – Six Sigma – Part 1: DMAIC methodology. Geneva: ISO copyright office, 2011. 32 s.
- [11] ISO 22514-3 Statistical methods in process management – Capability and performance – Part 3: Machine performance studies for measured data on discrete parts. Geneva: ISO copyright office, 2020. 19 s.
- [12] JAROŠOVÁ, E. – NOSKIEVIČOVÁ, D.: Pokročilejší metody statistické regulace procesu. Praha: Grada Publishing, 2015. 290 s. ISBN 978-80-247-5355-3.
- [13] JOGLEKAR, A. M.: Statistical Methods for Six Sigma in R&D and Manufacturing. Hoboken: J. Wiley and Sons, 2003. 321 s. ISBN 0-471-20342-4.
- [14] MILLER, I. – MILLER, M.: John E. Freund's Mathematical Statistics with Applications. Eighth Edition. Harlow: Pearson Education Limited, 2014. 472 s. ISBN 13: 978-1-292-02500-1.

- [15]MONTGOMERY, D. C.: Statistical Quality Control: A Modern Introduction. Seventh edition. Hoboken: J. Wiley and Sons, 2013. 754 s. ISBN 978-1-118-14681-1.
- [16]PILLET, M.: Six Sigma Comment l'appliquer. Deuxième tirage. Paris: Éditions d'Organisation, 2005. 486 s. ISBN 2-7081-3029-3.
- [17]PYZDEK, T.: The Six Sigma Handbook. Revised and Expanded. A Complete Guide for Green Belts, Black Belts, and Managers at All Levels. New York: McGraw-Hill, 2003. 830 s. ISBN 0-07-141596-3.
- [18]PYZDEK, T. – KELLER, P. A.: The Six Sigma handbook. A Complete Guide for Green Belts, Black Belts, and Managers at All Levels. Third Edition. New York: McGraw-Hill Companies, Inc., 2010. 548 s. ISBN 978-0-07-162337-7.
- [19]Six Sigma: Defects per Million Opportunities. 5 Steps for Calculating Defects per Million Opportunities (DPMO) [cit. 2023-05-16]. Dostupné na: <https://blog.masterofproject.com/DPMO-six-sigma/>.
- [20]STAMATIS, D. H.: Six Sigma for Financial Professionals. Hoboken: J. Wiley and Sons, 2003. 288 s. ISBN 0-471-45951-8.
- [21]TEREK, M. – HRNČIAROVÁ, Ľ.: Štatistické riadenie kvality. Bratislava: IURA EDITION, 2004. 234 s. ISBN 80-89047-97-1.
- [22]TEREK, M.: Interpretácia štatistiky a dát. 5., doplnené vydanie. Košice: Equilibria, 2017. 458 s. ISBN 978-80-8143-213-2.
- [23]TEREK, M.: Interpretácia štatistiky a dát. Podporný učebný materiál. 5., doplnené vydanie. Košice: Equilibria, 2017. 244 s. ISBN 978-80-8143-212-5.
- [24]TEREK, M.: Dotazníkové prieskumy a analýzy získaných dát. 1. vydanie. Košice: Equilibria, 2019. 200 s. ISBN 978-80-8143-247-7.
- [25]TEREK, M.: Metodológia určovania rozsahu výberového súboru. In: Slovenská štatistika a demografia, 2023. č. 1, s. 3 – 21.
- [26] What Is DPMO and How To Calculate It (With Examples) [cit. 2023-05-16]. Dostupné na: <https://www.indeed.com/career-advice/career-development/what-is-DPMO-and-how-to-calculate-it>.

RESUMÉ

V aplikáciách metodológie Six Sigma je pri hodnotení procesu nevyhnutné vhodne aplikovať charakteristiky výkonnosti procesu. Získané výsledky slúžia ako podklad na rozhodovanie o projektoch Six Sigma. Cieľom príspevku je podať súhrnný prehľad týchto charakteristík s dôrazom na ich štatistický význam, prepojenia medzi nimi a spôsoby získavania dát na ich výpočet. Vychádza sa z objasnenia štatistického významu Six Sigma a z modelu na meranie sigma úrovne kvality. Uvažuje sa o jedinej charakteristike kritickej pre kvalitu produktu, resp. o jednoduchom procese, ktorý sa skladá z jediného kroku. Predpokladá sa že má normálne rozdelenie. Keď je hodnota charakteristiky vnútri tolerančných hraníc, produkt je zhodný, keď nie, produkt je nezahodný. Pri návrhu koncepcie Six Sigma bol prijatý predpoklad, že stredná hodnota procesu je posunutá od cieľovej hodnoty o 1,5 smerodajnej odchýlky doprava alebo doľava. Pre rozličný počet smerodajných odchýlok charakteristiky vnútri tolerančných hraníc (sigma skóre) možno vypočítať napríklad počet nezahodných na milión jednotiek a iné ukazovatele. Ich hodnoty pre niektoré sigma skóre sú v tabuľke č. 1. Sigma úroveň kvality pre rozličný počet charakteristík kritických pre kvalitu možno odhadovať na základe ukazovateľa DPMO – počet nezahôd na milión príležitostí. Je uvedená tabuľka 2, z ktorej možno na základe vypočítanej hodnoty DPMO približne určiť sigma úroveň kvality (sigma skóre z). Uvádzajú sa aj dva alternatívne postupy na určenie sigma úrovne kvality na základe vypočítanej hodnoty DPMO.

Spôsobilosť a výkonnosť procesu možno spoľahlivo predvídať, len ak je proces stabilný, to znamená, že parametre rozdelenia pravdepodobnosti procesu sa v čase nemenia. Preto je dôležité zabezpečiť, aby sa pri odhadovaní DPMO vychádzalo zo stabilných procesov. Keď je proces regulovaný, možno realizovať jeden náhodný výber rozsahu n v čase, keď je proces stabilný. Aby sa zabezpečila platnosť podmienky náhodného vyberania z nekonečne veľkého základného súboru, že všetky jednotky sú z toho istého základného súboru, treba vybrať jednotky vyrobené v tom istom alebo približne v tom istom čase. Vzhľadom na to možno vybrať jednotky vyrobené v rade za sebou, teda v čo najkratšom čase. Podmienka nezávislosti by mala byť splnená tak, že jednotky sa vyrábajú nezávisle a teda výrobu každej jednotky možno považovať za realizáciu nezávislého náhodného pokusu. Rozsah výberu by mal byť aspoň 50 jednotiek, v prípade odhadovania DPMO pomocou intervalu spoľahlivosti, možno rozsah výberu pre určenú spoľahlivosť a najväčšiu chybu, vypočítať. Ďalšou podmienkou na určenie rozsahu výberu by malo byť splnenie požiadavky, aby sa na jednotkách náhodného výberu vyskytla aspoň jedna nezhoda. Všeobecne, pri implementácii náhodného vyberania z nekonečne veľkého základného súboru treba postupovať veľmi uvážlivo. Každý konkrétny prípad môže vyžadovať rozličnú procedúru vyberania. Keď je proces nestabilný, miery výkonnosti nemajú zmysluplnú interpretáciu, pretože nemôžu predikovať výkonnosť procesu. Často môže byť zaujímavý počet nezhôd na jednotku v základnom súbore, ktorý sa odhaduje z dát náhodného výberu hodnotou vypočítaného priemerného počtu nezhôd na jednu jednotku DPU. Celková výťažnosť RTY je pravdepodobnosť, že jednotka prejde všetkými krokmi procesu bez nezhôd.

RESUME

In the applications of the Six Sigma methodology, it is essential to appropriately apply process performance characteristics when evaluating a process. The obtained results serve as a basis for decision-making on Six Sigma projects. The aim of the paper is to provide an overview of these characteristics with an emphasis on their statistical significance, their interconnections, and data collection methods for their calculation. It is based on the clarification of the statistical meaning of Six Sigma and the model for measuring the sigma quality level. A single critical-to-quality characteristic is considered, or a simple process consisting of a single step. It is presumed to have a normal distribution. When the value of the characteristic is within the specification limits, the product is conforming, if not, the product is nonconforming. When designing the Six Sigma concept, an assumption was made that the process mean is shifted from the target value by 1,5 standard deviations to the right or to the left. For a different number of standard deviations of the characteristic within the specification limits (sigma quality level), for example, the number of nonconforming per million units and other indicators can be calculated. Their values for certain sigma quality level are provided in the Table No. 1. The sigma quality level for a different number of critical to quality characteristics can be estimated based on the DPMO indicator—the number of nonconformities per million opportunities. The table No. 2 is presented, from which the sigma quality level can be approximately determined based on the calculated DPMO value. Two alternative procedures for determining the sigma quality level based on the calculated DPMO value are also presented.

The capability and performance of a process can only be reliably predicted if the process is stable, that is, the parameters of the probability distribution of the process do not change over time. Therefore, it is important to ensure that the estimation of DPMO is based on stable processes. When the process is under control, one random

sample of size n can be selected at a time when the process is stable. To ensure the validity of the condition of random sampling from an infinite population i.e., that all units are from the same population, units produced at or around the same time should be selected. Therefore, sequentially produced units can be selected, i.e., in the shortest possible time. The condition of independence should be fulfilled in such a way that the units are produced independently and thus the production of each unit can be considered as an independent random experiment. The sample size should consist of at least 50 units, in case of the estimation of the DPMO by a confidence interval, the sample size can be calculated for a given confidence level and margin of error. Another condition for determining the sample size should be the fulfillment of the requirement that there is at least one nonconformity in the random sample units. In general, great care has to be taken when implementing random sampling from an infinite population. Each specific case may require a different sampling procedure. When a process is unstable, performance measures have no meaningful interpretation because they cannot predict process performance.

The number of nonconformities per unit in the population, which is estimated from random sample data by the calculated average number of nonconformities per unit DPU, might often be of interest. The rolled throughput yield RTY is the probability that a unit will pass through all process steps without nonconformities.

PROFESIJNÝ ŽIVOTOPIS

Prof. Ing. Milan Terek, PhD., od roku 2018 pracuje ako profesor na Vysokej škole manažmentu v Bratislave. Vede predmety úvod do štatistiky, štatistika, matematika pre manažérov II, kvantitatívne metódy pre manažérov a kvantitatívne metódy vo výskume v oblasti podnikového manažmentu. V rokoch 1977 – 2018 pracoval na Ekonomickej univerzite v Bratislave. Viedol predmety štatistika, štatistické riadenie kvality, analýza rozhodovania, hĺbková analýza dát, výberové skúmanie, lineárne programovanie, nelineárne programovanie, operačný výskum a systémové modelovanie. Vo výskume sa zameriava na aplikácie štatistických metód v ekonómii a manažmente. Je autorom alebo spoluautorom 6 monografií, 10 vysokoškolských učebníc, 17 skrípt, 78 článkov vo vedeckých a odborných časopisoch a 115 príspevkov na vedeckých konferenciách publikovaných v zborníkoch.

KONTAKT

mterek@vsm.sk

Informatívny článok/Informative article

Roman PAVELKA
Štatistický úrad Slovenskej republiky

ÚVOD DO STATISTICKÉHO MODELOVÁNÍ POMOCÍ ANALYTICKÉHO SYSTÉMU SAS

INTRODUCTION TO STATISTICAL MODELLING USING THE SAS ANALYSIS SYSTEM

ABSTRAKT

Mohutný nárůst výpočetního výkonu ve druhé polovině 20. století a vznik nových moderních technologií na počátku 21. století podstatně ovlivnil rozvoj metodologie statistického modelování. Od skromných začátků v zemědělství se aplikace statistického modelování, a to zejména lineární modely, staly zásadními v mnoha oblastech vědy a výzkumu. V průběhu několika desítek let se metodologie statistických modelů přetvořila do obecných analytických postupů datové vědy a tvoří jádro moderních statisticko-analytických metod. V popředí vývoje a implementace uvedených metod stojí programový systém SAS, který již ve svých základních modulech umožňuje tyto poznatky využívat v reálné praxi. I když je systém SAS vybaven mnohými funkcionalitami, které odpovídají současnému stavu statistické vědy, není v našich podmínkách příliš rozšířen, a to především z důvodu vyšší pořizovací ceny. Proto je cílem tohoto článku přiblížit některé možnosti tohoto softvéru z oblasti statistického modelování, které zůstávají mnohým analytikům skryté.

ABSTRACT

The massive increase in computing power in the second half of the 20th century and the emergence of new modern technologies in the early 21st century have significantly influenced the development of statistical modelling methodology. From humble beginnings in agriculture, applications of statistical modelling, and in particular linear models, have become fundamental in many areas of science and research. Over the course of several decades, statistical model methodology has evolved into general data science analytical procedures and forms the core of modern statistical analytical methods. At the forefront of the development and implementation of these methods is the SAS software system, which enables these findings to be used in real practice already in its basic modules. Although the SAS system is equipped with many functionalities that correspond to the current state of statistical science, it is not very widespread in our conditions, mainly due to its higher purchase price. Therefore, the aim of this article is to present some of the capabilities of this software in the field of statistic modelling, which remain hidden to many analysts.

KLÍČOVÉ SLOVA

analýza rozptylu, regresní analýza, SAS, statistické modelování, testování hypotéz

KEY WORDS

analysis of variance, regression analysis, SAS, statistical modelling, hypothesis testing

1. ÚVOD

Programový systém SAS vznikl v roce 1976 pod označením SAS-76. Již od počátku existence systému SAS se vlajkovou lodí statistického modelování stala procedura GLM. Na svou dobu byla tato procedura velmi inovativní a upoutala pozornost statistiků a dalších osob zabývajících se analýzou dat v USA i mimo ně. Procedura GLM poskytovala komplexní platformu, která umožňovala práci s různými typy lineárních modelů. Uživatelům umožnila získat řešení pro většinu problémů spadajících do oblasti lineárních modelů, pro regresní analýzu, analýzu lineárních modelů a analýzu rozptylu i kovariance a analýzu vícerozměrných dat. Procedura GLM se stala vzorem pro vývoj další statistické procedury v systému SAS¹ a pro většinu statistiků a datových analytiků základním postupem při práci s daty [4].

V průběhu následujících let byla vytvořena statistická procedura REG, která rozšířila možnosti regresní analýzy o diagnostické nástroje. Nyní měl uživatel nejen možnost odhadovat inferenční statistiky v regresní analýze, ale mohl také získat statistiky, které mu pomohou rozhodnout, jaké proměnné do analýzy zahrnout, a identifikovat problematické data [1]. Postupující pokrok výpočetní techniky přinesl překotný rozvoj ve vývoji statistických procedur systému SAS. Možnosti statistického modelování byly rozšířeny na nelineární modely, modely s pevnými i náhodnými efekty i na problémy spojené s analýzou korelovaných dat. V prvních verzích systému SAS byly značně omezené možnosti analýzy kategoriálních dat. Proto byl tento analytický softvér obohacen o statistické procedury, které inovovaly použití lineárních modelů pro analýzu kategoriálních dat. Toto vylepšení proto umožňovalo řešení problémů zobecněných lineárních modelů [6]. V současnosti systém SAS svým uživatelům nabízí v modulu SAS/STAT několik desítek procedur k analýze dat [8], která pokrývají většinu problémů moderní statistiky včetně vybraných oblastí strojového učení a umělé inteligence. Systém SAS je navíc průběžně aktualizován o nejnovější postupy datové vědy.

2. STATISTICKÉ MODELOVÁNÍ V PROSTŘEDÍ SYSTÉMU SAS

2.1. ZÁKLADNÍ POJMY A DEFINICE

Statistický model [9] je stochastický model, který obsahuje parametry (neznámé konstanty), které je třeba odhadnout na základě předpokladů o modelu a pozorovaných datech. Statistické modely mohou být jednoduché, ale také i velmi komplexní. Proto je nejvhodnější klasifikovat modely podle jednoduchých kritérií, jako je např. přítomnost náhodných efektů, přítomnost nelinearity, charakteristiky dat atd. Statistický model popisuje distribuční vlastnosti jedné nebo více vysvětlovaných proměnných. Tento popis má často jednoduchou podobu modelu s aditivní chybovou strukturou:

$$\text{Vysvětlovaná proměnná} = \text{průměr} + \text{chyba}$$

V matematickém zápisu má tato jednoduchá rovnice modelu tvar:

$$Y = f(x_1, \dots, x_k; \beta_1, \dots, \beta_p) + \varepsilon \quad (1)$$

¹ *Statistický systém SAS (z angl. Statistical Analysis System) je integrovaný systém softvérových produktů pro analýzu dat vyráběný americkou firmou SAS Institute, Inc.*

V rovnici (1) je proměnná Y nazývaná závislou, vysvětlovanou nebo výstupní proměnnou. Výrazy x_1, \dots, x_k označují hodnoty k regresorových proměnných, také označovaných jako kovariáty, nezávislé nebo vysvětlující proměnné. Neznámé konstanty modelu, které jsou předmětem statistického odhadu, jsou označeny jako β_1, \dots, β_p . Člen rovnice (1) označený ε vyjadřuje náhodné poruchy modelu; nazývá se také reziduální nebo chybový člen modelu.

Je-li v rovnici modelu (1) zdrojem náhodnosti pouze chybový člen ε a jestliže mají tyto náhodné chyby nulový průměr, potom funkce (1) se stává funkcí střední hodnoty daného statistického modelu ve tvaru

$$E[Y] = f(x_1, \dots, x_k; \beta_1, \dots, \beta_p) \quad (2)$$

kde $E[\cdot]$ označuje operátor střední hodnoty (očekávání).

V mnoha aplikacích je jednoduchá formulace modelu nedostatečná. Často je potřebné specifikovat nejen stochastické vlastnosti jednoduchého chybového členu, ale také to, jak se chyby modelu spojené s různými pozorováními navzájem ovlivňují. Pokud chyby nemají nulovou střední hodnotu nebo pokud rozptyl pozorování závisí na podmíněných středních hodnotách, k popisu mechanismu generujícího data jednoduchý aditivní chybový model je obvykle nedostatečný. Modely pro taková data obvykle vyžadují složitější formulace zahrnující různá rozdělení pravděpodobnosti.

2.2. TRÍDY STATISTICKÝCH MODELŮ DOSTUPNÉ V SYSTÉMU SAS

Lineární a nelineární modely

Problém statistického odhadu je nelineární, pokud odhadové rovnice – rovnice, jejichž řešení dávají odhady parametrů, závisí na parametrech nelineárně. Takové problémy odhadů obvykle nemají řešení v uzavřené formě a musí se řešit iteračními numerickými technikami.

K rozlišení mezi lineárními a nelineárními modely se často používá nelinearita funkce střední hodnoty. Model má nelineární funkci střední hodnoty, jestliže derivace funkce střední hodnoty vzhledem k jejím parametrům závisí alespoň na jednom dalším parametru. Nelineární funkce střední hodnoty vede k nelineárním odhadům.

Jednorozměrné a vícerozměrné modely

Vícerozměrný statistický model je model, ve kterém je více vysvětlovaných proměnných modelováno současně. Například pokud data obsahují hodnoty výšky (h_i) a váhy (w_i) dětí sbíraných za několik let, potom vícerozměrný statistický model je dán výrazem:

$$Y_i = \begin{bmatrix} w_i \\ h_i \end{bmatrix} = X\beta + \begin{bmatrix} \varepsilon_{wi} \\ \varepsilon_{hi} \end{bmatrix} = X\beta + \varepsilon_i \quad (3)$$

Vektory vysvětlovaných proměnných Y_i a náhodných chyb ε_i obsahují po 2 pozorování patřících i -tému dítěti. Chyby téhož dítěte mají tak korelaci:

$$CORR[\varepsilon_{wi}, \varepsilon_{hi}] = \frac{\sigma_{wh}}{\sqrt{\sigma_w^2 \sigma_h^2}} \quad \varepsilon_i = \left(\mathbf{0}, \begin{bmatrix} \sigma_w^2 & \sigma_{wh} \\ \sigma_{wh} & \sigma_h^2 \end{bmatrix} \right), \quad (4)$$

kde σ_w^2 označuje rozptyl hodnot váhy dětí, σ_h^2 je rozptyl hodnot výšky dětí a σ_{wh} jejich kovariance.

Regresní modely a modely s klasifikačními efekty

Regresní model v užším slova smyslu v porovnání s klasifikačním modelem je modelem lineárním, pokud jsou všechny regresorové proměnné (efekty) v modelu spojité. Jinak řečeno, každý spojitý efekt v modelu přispívá jedním sloupcem do matice modelu \mathbf{X} , a tedy jedním parametrem do modelu jako celku.

Klasifikační efekt je naopak spojen s více než jedním sloupcem matice \mathbf{X} . Klasifikace vzhledem k proměnné je proces, při kterém je každé pozorování přiřazeno jednomu z k úrovní; proces určení těchto k úrovní se označuje jako tzv. levelizace proměnné [9]. Klasifikace proměnných se v modelech používá k určení experimentálních podmínek, příslušnosti ke skupině, ošetření atd. a podobně. Skutečné hodnoty klasifikační proměnné nejsou důležité a proměnná může být číselná nebo číselně vyjádřená znaková proměnná. Důležitá je asociace diskretních hodnot nebo úrovní klasifikační proměnné se skupinami pozorování. Příkladem modelu s klasifikačním efektem je model, ve kterém je zahrnuta dvojúrovňová klasifikační proměnná GENDER. Hodnoty této klasifikační proměnné jsou kódovány jako 'F' a 'M'. Potom symbolické vyjádření modelu ve tvaru

$$weight = age + bmi + gender + error \quad (5)$$

se rozšiřuje do statistického modelu

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \tau_1 I(gender = 'F') + \tau_2 I(gender = 'M') + \varepsilon_i \quad (6)$$

kde $I(gender='F')$ je indikátorová funkce, která vrací 1, pokud je hodnota proměnné pohlaví 'F', a 0 v opačném případě. Parametry τ_1 a τ_2 jsou spojeny s efektem klasifikace proměnné GENDER. Tato forma zahrnutí klasifikační proměnné do modelu je pouze jedním z několika různých způsobů, jak do modelu zahrnout vliv úrovní klasifikační proměnné. Jedná se o tzv. singulární parametrizaci², která je obecně používaným přístupem a používá se v procedurách GLM, MIXED a GLIMMIX. Jiné tvary parametrizace klasifikačních efektů pomocí různých forem tzv. nesingulární parametrizace jsou k dispozici v procedurách GENMOD a LOGISTIC.

Modely, které obsahují pouze klasifikační efekty, se často ztotožňují s modely analýzy rozptylu (ANOVA), protože při jejich analýze se často používají metody ANOVA. To platí zejména pro experimentální data, kde modelové efekty zahrnují efekty ošetření a návrhu řízení chyb a pod. Nicméně klasifikační proměnné se objevují

² Provedenou parametrizací se stanou sloupce matice \mathbf{X} lineárně závislými, matice $\mathbf{X}'\mathbf{X}$ je singulární a pro řešení normálních rovnic je nutné použít zobecněnou inverzní matici $\mathbf{X}'\mathbf{X}$ podle (Ben-Israel, Greville, 2003).

ve větší míře i v jiných modelech. Například mnoho smíšených modelů, kde se parametry odhadují pomocí maximální věrohodnosti, se skládá výhradně z klasifikačních efektů. Tyto modely ale neumožňují rozklad součtu čtverců typický pro techniku ANOVA. Mnoho modelů obsahuje jak spojité, tak klasifikační efekty. Například statistický model se spojitou proměnnou v jednotlivých úrovních klasifikačního efektu. Takové efekty jsou vhodné například k tomu, aby se v regresním modelu měnily sklony podle úrovní klasifikační proměnné.

Modely s pevnými efekty, s náhodnými efekty a modely se smíšenými efekty

Ve statistickém modelu představuje každá nezávislá proměnná pevný nebo náhodný efekt. Pevné efekty v modelu jsou neznámé konstanty (parametry). Naopak náhodné efekty konstantní nejsou a podléhají normálnímu rozdělení. Modely, ve kterých jsou všechny efekty pevné, se nazývají modely s pevnými efekty. Podobně modely, v nichž jsou všechny efekty náhodné – kromě případného absolutního členu (intercept) – se nazývají modely s náhodnými efekty. Smíšené modely jsou tedy takové modely, které obsahují pevné a náhodné efekty. V maticovém zápisu lineární model s efekty pevnými, s efekty náhodnými nebo lineární model se smíšenými efekty představují následující modelové rovnice:

$$Y = X\beta + \varepsilon \quad (7)$$

$$Y = Z\gamma + \varepsilon \quad (8)$$

$$Y = X\beta + Z\gamma + \varepsilon \quad (9)$$

V rovnicích (7) až (9) jsou X , resp. Z regresní matice modelu spojené s pevnými, resp. náhodnými efekty. Vektor β je vektor parametrů pevných efektů a vektor γ reprezentuje náhodné efekty. Procedury pro modelování se smíšenými modely v softvéru SAS/STAT předpokládají, že náhodné efekty sledují normální rozdělení s rozptylovou a kovarianční maticí G a ve většině případů mají náhodné efekty nulovou střední hodnotu.

Zobecněné lineární modely

Třídou modelů, která v posledních desetiletích nabývá na významu, je třída zobecněných lineárních modelů. Teorie zobecněných lineárních modelů vychází z odborných statí [6, 10], následně byla zpopularizována v monografii [5].

Toto zobecnění vyžaduje podstatně složitější nastavení statistického modelu, než je tomu v případě lineárních modelů s normálně rozdělenými daty. Zobecněný lineární model se skládá ze:

- **systematické složky**, která v modelu vystupuje jako lineární prediktor podobně jako v lineárních modelech. Lineární prediktor $\eta = x'\beta$ je lineární funkce v parametrech. Na rozdíl od lineárního modelu nepředstavuje tzv. funkci střední hodnoty podle (2).
- **spojovací funkce** $g(\mu) = \eta$, což je funkce, která spojuje lineární prediktor s celkovým průměrem μ . Spojovací funkce je monotónní, inverzní funkce. Střední hodnotu lze tedy vyjádřit jako inverzní lineární prediktor, $\mu = g^{-1}(\eta)$. Například běžnou spojovací funkcí pro binární a binomická data je logitová spojovací funkce, $g(t) = \ln\{t/(1-t)\}$.

- **náhodná složka** zobecněného lineárního modelu, která je reprezentována rozdělením pravděpodobnosti. Předpokládá se, že rozdělení pravděpodobnosti náleží do skupiny exponenciálních. Mezi diskrétní rozdělení této rodiny patří např. Bernoulliho (binární), binomické, Poissonovo, geometrické a záporné binomické rozdělení (pro danou hodnotu škály). Ke spojitým rozdělením pravděpodobnosti patří normální (Gaussovo), beta, gama, inverzní Gaussovo a exponenciální rozdělení.

Speciálním případem zobecněného lineárního modelu je klasický lineární model, kde spojovací funkce je funkcí identity a rozdělení pravděpodobnosti je normální.

Programový systém SAS pro odhad parametrů statistických modelů používá klasické odhadové metody, a to metodu nejmenších čtverců (včetně vážených, s iteračními výpočty) a metody založené na věrohodnostní funkci. Možnosti statistické inference nad daty ze statistických zjišťování poskytují procedury SURVEYMEANS, SURVEYFREQ, SURVEYREG, SURVEYLOGISTIC a SURVEYPHREG, které při odhadech dovolují zohlednit také i mechanismus výběru do zjišťování.

3. ZÁKLADNÍ PRINCIPY REGRESNÍ ANALÝZY V SYSTÉMU SAS

Programový systém SAS dokáže analyzovat široké spektrum regresních modelů. Statistické procedury, kterými je SAS vybaven, podporují regresní modely:

- s diskrétní nebo spojitou vysvětlovanou proměnnou, jejíž rozdělení může, ale také nemusí podléhat normálnímu rozdělení pravděpodobnosti,
- lineární, nelineární anebo zobecněné lineární formy statistických modelů,
- modely obsahující klasifikační proměnné (nevstupují do modelu prostřednictvím svých hodnot, ale prostřednictvím svých úrovní),
- u modelů s klasifikačními efekty se pro zajištění jednoznačnosti odhadů využívá konceptu tzv. odhadnutelných funkcí,
- statistické modely s daty generovanými z experimentů, pozorování anebo statistických šetření,
- pro metody odhadů pomocí metod založených na funkci maximální věrohodnosti a minimalizaci součtu reziduálních čtverců.

3.1. ODHADY PARAMETRŮ LINEÁRNÍHO MODELU V SYSTÉMU SAS

Odhady parametrů modelu β metodou nejmenších čtverců (příp. vážených W) jsou dány řešením normálních rovnic:

$$(X'WX)\beta = X'WY \quad (10)$$

Jediným předpokladem, který je nutný k tomu, aby odhady metodou nejmenších čtverců byly nestranné, je nulová střední hodnota chyb modelu. Získané odhady

$$\hat{\beta} = (X'WX)^{-1}X'WY \quad (11)$$

mají minimální rozptyl ve třídě odhadů, které jsou nestranné a jsou lineárními funkcemi vysvětlované proměnné. Pokud je splněn další předpoklad normálně rozdělených chyb, pak platí následující [9]:

- vypočtené statistiky mají výběrová rozdělení vhodná pro testování hypotéz,
- odhady parametrů jsou normálně rozděleny,
- různé součty čtverců jsou alespoň za platnosti nulových hypotéz rozděleny úměrně chí-kvadrát rozdělení,

- poměry odhadů ke standardním chybám se za platnosti nulových hypotéz řídí Studentovým t rozdělením,
- poměry středních čtverců se řídí F rozdělením za platnosti nulových hypotéz.

Pokud jsou ke statistickému modelování použita data, která nespĺňují výše uvedené předpoklady normality rozdělení chyb, výsledky analýzy je nutné interpretovat opatrně. Pravděpodobnosti významnosti jsou za těchto okolností nespolehlivé.

Pro realizaci odhadů lineárního modelu s funkcí střední hodnoty $E[Y] = \beta X$ metodou maximální věrohodnosti je potřebné specifikovat rozdělení pravděpodobnosti vysvětlované proměnné Y . Pro odhady parametrů lineárního modelu lze podle [8] použít věrohodnostní funkci (resp. její logaritmus) ve tvaru

$$l(\beta, \sigma^2; y) = -\frac{n}{2} \ln\{2\pi\} - \frac{n}{2} \ln\{\sigma^2\} - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \quad (12)$$

Funkce (12) je maximalizována, když je minimalizován součet čtverců $(y - X\beta)'(y - X\beta)$. Odhad parametrů β v lineárním modelu metodou maximální věrohodnosti je proto totožný s odhadem metodou nejmenších čtverců. Pokud matice $X'WX$ nemá plnou hodnotu, je nutné k řešení normálních rovnic pro minimalizaci součtu čtverců $(y - X\beta)'(y - X\beta)$ použít metodu zobecněné inverzní matice [2].

Pokud modelová matice X obsahuje sloupec, jehož nenulové hodnoty se nemění, obvykle sloupec jedniček, obsahuje lineární model absolutní člen (intercept). Možnost potlačit automatické přidání sloupce jedniček umožňují statistické procedury, které podporují parametr NOINT v příkazu MODEL. Obecně by modely bez absolutního členu (interceptu) měly být výjimkou, zejména pokud navrhovaný model neobsahuje klasifikační proměnné.

Vyrovnané hodnoty \hat{y}_i pro $i = 1 \dots n$ jsou reprezentovány rovnicí:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k \quad (13)$$

Po provedeném odhadu parametrů chybová složka pro i -tou hodnotu je dána jako

$$\hat{\varepsilon}_i = y_i - \hat{y}_i. \quad (14)$$

Tabulka č. 1: Analýza rozptylu (ANOVA) lineárního modelu

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	k	$SS_M = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$MS_M = \frac{SS_M}{k}$	$F = \frac{MS_M}{MS_E}$	Hodnota p-value <
Error	n - k - 1	$SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$MS_E = \frac{SS_E}{n - k - 1}$		
Corrected Total	n - 1	$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2$			

Poznámka: Symbol n vyjadřuje počet pozorování a $k+1$ vyjadřuje počet parametrů v modelu, z toho k je počet vysvětlujících proměnných.

Zdroj: vlastní zpracování autora podle [9]

Současně s odhady parametrů regresní procedurou v systému SAS se analýzou rozptylu testuje významnost odhadnutého modelu (tzv. celkový F -test). Vyšší hodnota pravděpodobnosti (p -value $> 0,05$) nasvědčuje nevýznamnosti modelu. Analýza rozptylu s rozkladem sum čtverců a testové F statistiky jsou ilustrovány v tabulce č. 1.

3.2. ODHADY PARAMETRŮ OBECNÉHO LINEÁRNÍHO MODELU POMOCÍ SAS

Obecný lineární model lze podobně jako lineární model (regresní model v užším významu) vyjádřit v podobě modelu s aditivní chybovou složkou [7]. Na rozdíl od klasického lineárního modelu je koncepce obecného lineárního modelu založena na podstatně obecnějších základech. Původní chápání statistického modelu s aditivní strukturou chyby (vyrovnaná + reziduální složka) bylo nahrazeno koncepcí symbolicky vyjádřenou ve tvaru:

$$\mathbf{Data} = \mathbf{Model} + \mathbf{Error} \quad (15)$$

V rovnici (15) je model vyjádřením pochopení podstaty a mechanismu vzniku dat, resp. hypotézy o nich, které jsou aplikovány na analyzovaná data. Chybová složka označená *Error* je explicitním vyjádřením toho, že na data působí i jiné vlivy a okolnosti než ty zahrnuté v modelu. V rámci procesu obecného lineárního modelování se analytici pokoušejí porovnáním různých lineárních modelů určit *nejlepší* obecný lineární model pro data. Testované obecné lineární modely jsou posuzovány z hlediska relativního podílu rozptylu dat připisovaného modelu a chybových složek. U neměnného datového souboru je součet modelové a chybové složky variance (tj. rozptyl dat) konstantní, takže každé zvýšení variance, které je vysvětleno modelovou složkou, bude mít za následek ekvivalentní snížení variance chybové složky. Výraz *obecný* v modelu jednoduše označuje schopnost zohlednit rozdíly kvantitativních proměnných, které představují spojitě míry (jako v regresi) a diskrétní rozlišení (kategoriální proměnné) reprezentující skupiny nebo experimentální podmínky. Podobně jako v analýze kovariance jsou v obecném lineárním modelu obsaženy proměnné reprezentující jak kvantitativní veličiny, tak kategoriální úrovně. Podobně jako v analýze rozptylu jsou porovnávány rozdíly mezi průměrnými hodnotami závislé (vysvětlované) proměnné na jednotlivých úrovních kategoriální proměnné (nazývané faktorem) a je testována významnost rozdílu uvedených průměrů. V obecném lineárním modelu může být zahrnut 1 nebo více různě uspořádaných faktorů (např. znáhodněné bloky, latinské čtverce atd.), opakovaná měření s vyváženými či nevyváženými návrhy. Proto regresní analýza (lineární modelování), analýza rozptylu a analýza kovariance jsou pouze speciálními případy obecných lineárních modelů.

Pro ukázkou možné analýzy v prostředí SAS lze obecný lineární model formulovat ve tvaru

$$y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \varepsilon_{ijk}, \quad (16)$$

kde y_{ijk} je k -té pozorování vysvětlované (závislé) na i -té úrovni faktoru A a j -té úrovni faktoru B ,
 μ je absolutní člen (intercept),
 A_i je i -tá úroveň faktoru A ,
 B_j je j -tá úroveň faktoru B
 $(AB)_{ij}$ je interakce mezi hodnotou i -té úrovně faktoru A a j -té úrovni faktoru B a
 ε_{ijk} je k -tá reziduální hodnota i -té úrovně faktoru A na j -té úrovni faktoru B .

Náhodné chyby ε_{ijk} jsou nezávislé a stejně rozdělené s rozdělením $N(0, \sigma_E^2)$.³

Tvar modelu podle (16) popisuje obecný lineární model se 2 faktory (označené A a B) s jejich vzájemnou interakcí $(AB)_{ij}$. Pro účely analýzy rozptylu (porovnání průměrů na různých úrovních obou faktorů) je vhodný model (16) přepsat do tvaru

$$y_{ijk} = \mu + \underbrace{\mu_i - \mu}_{\text{efekt faktoru A}} + \underbrace{\mu_j - \mu}_{\text{efekt faktoru B}} + \underbrace{\mu_{ij} - \mu_i - \mu_j + \mu}_{\text{efekt interakce}} + \varepsilon_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \quad (17)$$

kde y_{ijk} je k -té pozorování i -té úrovně faktoru A na j -té úrovni faktoru B ,
 μ je absolutní člen (intercept),
 $(\mu_i - \mu)$ je efekt faktoru A , kde μ_i je marginální střední hodnota i -té úrovně faktoru A ,
 $(\mu_j - \mu)$ je efekt faktoru B , kde μ_j je marginální střední hodnota j -té úrovně faktoru B ,
 $(\mu_{ij} - \mu_i - \mu_j + \mu)$ je efekt interakce i -té úrovně faktoru A a j -té úrovně B ,
 indexy obsahující symbol \cdot označují průměrné hodnoty na dané úrovni.

Programový systém SAS je k analýze rozptylu vybaven několika statistickými procedurami. Pro zpracování vyvážených dat (tj. dat se stejným počtem pozorování pro každou kombinaci klasifikačních faktorů) je určena především procedura ANOVA. Pokud každá kombinace úrovní klasifikačních faktorů neobsahuje stejný počet pozorování (tj. jedná se o nevyvážená data), musí se použít procedura GLM. Procedura GLM⁴ poskytuje analýzu vyvážených i nevyvážených dat s možností odhadu parametrů obecného lineárního modelu [9].

Tabulka č. 2: Analýza rozptylu (ANOVA) obecného lineárního modelu se 2 faktory

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Faktor A	$n_A - 1$	$SS_A = n_B * n_I * \sum_{i=1}^{n_A} (\mu_i - \mu)^2$	$MS_A = \frac{SS_A}{n_A - 1}$	$F_A = \frac{MS_A}{MS_E}$	Hodnota p-value <
Faktor B	$n_B - 1$	$SS_B = n_A * n_I * \sum_{j=1}^{n_B} (\mu_j - \mu)^2$	$MS_B = \frac{SS_B}{n_B - 1}$	$F_B = \frac{MS_B}{MS_E}$	Hodnota p-value <
Interakce A*B	$(n_A - 1) * (n_B - 1)$	$SS_{AB} = n_I * \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} (\mu_{ij} - \mu_i - \mu_j + \mu)^2$	$MS_{AB} = \frac{SS_{AB}}{(n_A - 1) * (n_B - 1)}$	$F_{AB} = \frac{MS_{AB}}{MS_E}$	Hodnota p-value <
Error	$n_A * n_B * (n_I - 1)$	$SS_E = \sum_{k=1}^{n_I} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} (y_{ijk} - \mu_{ij})^2$	$MS_E = \frac{SS_E}{n_A * n_B * (n_I - 1)}$		
Corrected Total	$n_A * n_B * n_I - 1$	$SS_T = \sum_{k=1}^{n_I} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} (y_{ijk} - \mu)^2$			

Poznámka: „ n_A “ označuje počet pozorování pro faktor A „ n_B “ označuje počet pozorování pro faktor B
 „ n_I “ označuje počet pozorování interakce faktorů A*B „ \cdot “ označují průměrné hodnoty na dané úrovni

Zdroj: vlastní zpracování autora podle [9]

Výsledkem analýzy rozptylu procedury GLM je tabulka analýzy rozptylu s relativními podíly rozptylu připadajícími na vliv jednotlivých faktorů a jejich interakce na rozptylu celkovém. Jednotlivé podíly představují statistiky F určené k testování významnosti faktorů a jejich interakce v modelu. Vysoká hodnota pravděpodobnosti nasvědčuje platnosti nulové hypotézy o nevýznamnosti modelu. Pokud je toto číslo nižší než 5 %

³ Normální rozdělení náhodných chyb (identicky a vzájemně nezávislé hodnoty) je hlavním rozdílem od zobecněných lineárních modelů, ve kterých tento předpoklad nemusí platit.

⁴ GLM znamená General Linear Model tj. v překladu obecný lineární model.

(0,05), je obvykle vliv faktoru, resp. interakce považován za dostatečně významný. Definice analýzy rozptylu pro dvoufaktorový model s interakcemi je v tabulce č. 2.

Odhad parametrů obecného lineárního modelu (16) je realizován procedurou GLM metodou nejmenších čtverců řešením normálních rovnic (10). Faktory jako kategorické proměnné se do matice modelu \mathbf{X} zahrnou pomocí tzv. singulární parametrizace (6). Na základě toho je možné obecný lineární model přeformulovat do podoby vícenásobného regresního modelu, na jehož základě procedura GLM odhaduje parametry modelu:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (18)$$

Parametry regresního modelu (18) jsou odhadnuty v souladu s rovnicí (11). Odhadnuté parametry $\hat{\beta}$ v důsledku singulární parametrizace (6) však nepředstavují jedinečné řešení. Pro zajištění jednoznačnosti odhadu $\hat{\beta}$ se využívají tzv. odhadnutelné funkce $\mathbf{L}\beta$ [3]. Základem odhadnutelných funkcí $\mathbf{L}\beta$ je vektor prvků \mathbf{L} , které reprezentují váhy odpovídající každému efektu, případně interakci efektů ve funkci odhadů středních hodnot. Počet jedinečných symbolů ve vektoru představuje maximální počet lineárně nezávislých koeficientů odhadovaných modelem, který je roven hodnotě matice $\mathbf{X}'\mathbf{X}$. Odhadnutelné funkce $\mathbf{L}\beta$ se vyznačují následujícími vlastnostmi:

- $\widehat{\mathbf{L}\beta}$ a kovarianční matice $\text{VAR}(\widehat{\mathbf{L}\beta})$ jsou jedinečné,
- $\widehat{\mathbf{L}\beta}$ je jedinečným odhadem $\mathbf{L}\beta$, jejichž kovarianční matice je dána výrazem:

$$\text{VAR}(\widehat{\mathbf{L}\beta}) = \sigma_\varepsilon^2 [\mathbf{L}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{L}'], \quad (19)$$

kde \mathbf{L} je vektor prvků reprezentující váhy odpovídající každému efektu,
 \mathbf{X} je matice modelu,
 \mathbf{W} je matice vah (v případě klasického lineárního modelu je jednotková) a
 σ_ε^2 je rozptyl náhodné složky v modelu.

Na základě modelové složky dvoufaktorové rovnice obecného lineárního modelu jsou předpovídány hodnoty dány následujícím vztahem [7]:

$$\hat{y}_{ijk} = \mu + \underbrace{\mu_{i.} - \mu}_{\text{efekt } i} + \underbrace{\mu_{.j} - \mu}_{\text{efekt } j} + \underbrace{\mu_{ij} - \mu_{i.} - \mu_{.j} + \mu}_{\text{interakce } ij} \quad (20)$$

Rozdíly mezi hodnotami pozorovanými y_{ijk} a vyrovnanými \hat{y}_{ijk} modelovanými procedurou GLM jsou definovány výrazem

$$\hat{\varepsilon}_{ijk} = y_{ijk} - \hat{y}_{ijk}. \quad (21)$$

Pomocí příkazů procedury GLM lze odhadovat také marginální střední hodnoty jednotlivých faktorů (příkaz LSMEANS) a jejich interakce, průměry na jednotlivých úrovních faktorů a interakce (s využitím odhadnutelných funkcí \mathbf{L}) a další statistiky.

Odhady lineárních funkcí parametrů lineárního modelu

Procedura GLM jako jedna z nejstarších statistických procedur v systému SAS dokáže nejen odhadovat samotné parametry lineárních modelů, ale je vybavena také příkazy a parametry k odhadům lineárních funkcí parametrů modelu. Odhady

lineárních funkcí parametrů modelu vytváří podmínky pro testování lineárních modelů pomocí obecných lineárních hypotéz [4, 10], které tak mohou být konstruovány od jednoduchých až po komplexní porovnávání. Pro generování výstupů s lineárními kombinacemi odhadů parametrů modelu pro testování obecnými lineárními hypotézami je procedura GLM vybavena příkazy ESTIMATE a CONTRAST⁵.

Podobně jako hypotéza o nevýznamnosti odhadnutého modelu (celkový F -test podle tabulky č. 1) nebo analýza rozptylu obecného lineárního modelu (viz tabulka č. 2) je obecná lineární hypotéza [4] založena na vhodném rozkladu součtu čtverců a testování relativního podílu rozptylu připadajícího na odhadovanou funkci parametrů modelu (F statistika) [7].

Základní syntaxe příkazu ESTIMATE je:

```
estimate 'label' effect-name effect-coefficients;
```

kde koeficienty efektů jsou reprezentovány lineární kombinací parametrů modelu a 'label' představuje znakový řetězec identifikující výsledek ve výstupu procedury GLM.

Nechť má obecný lineární model (16) 2 pevné faktory, a to faktor A se 3 úrovněmi a B se 2 úrovněmi. Potom lze obecný lineární model podle (16) vyjádřit rovnicí jako:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad (24)$$

kde $i = 1, 2, 3$, $j = 1, 2$ a $k = 1, \dots, n_{ij}$.

Vektor odhadovaných parametrů β odpovídajících modelu podle vztahu (24) je:

$$\beta = (\mu \ \alpha_1 \ \alpha_2 \ \alpha_3 \ \beta_1 \ \beta_2 \ (\alpha\beta)_{11} \ (\alpha\beta)_{12} \ (\alpha\beta)_{21} \ (\alpha\beta)_{22} \ (\alpha\beta)_{31} \ (\alpha\beta)_{32}) \quad (25)$$

Vektor parametrů (25) se z pohledu jednotlivých faktorů dá zjednodušeně vyjádřit:

$$\beta = (\text{Intercept} | \text{Faktor A} | \text{Faktor B} | \text{Interakce}) \quad (26)$$

Odhadovaná lineární kombinace parametrů se definuje jako odhadnutelná funkce $L\beta$ [3], ve které je vektor L konstruován podobně jako vektor parametrů modelu s vahami odpovídajícími každému efektu v modelu. Formálně lze vektor vah L vyjádřit

$$L = \left(\begin{array}{c|c|c|c} \text{Váha pro} & \text{Váha pro} & \text{Váha pro} & \text{Váha pro} \\ \text{Intercept} & \text{Faktor A} & \text{Faktor B} & \text{Interakci} \end{array} \right) \quad (27)$$

Například funkce střední hodnoty vycházející z rovnice (24) určená k odhadu střední hodnoty $\mu_{1.}$ úrovně 1 faktoru A, tj. A_1 , se vyjádří ve tvaru:

$$\mu_{1.} = \mu + \alpha_1 + \frac{\beta_1}{2} + \frac{\beta_2}{2} + \frac{(\alpha\beta)_{11}}{2} + \frac{(\alpha\beta)_{12}}{2}. \quad (28)$$

⁵ Příkazy ESTIMATE a CONTRAST disponují také další procedury lineárního modelování jako je procedura GENMOD, GLIMMIX, LOGISTIC, MIXED, SURVEYREG a SURVEYLOGISTIC.

Vektor vah L , který odpovídá funkci odhadu podle (28), je definován jako:

$$L = (1 \mid 1 \ 0 \ 0 \mid 0.5 \ 0.5 \mid 0.5 \ 0.5 \ 0 \ 0 \ 0 \ 0) \quad (29)$$

Váhy ve vektoru L musí být podle [7] pro každý faktor rozděleny rovnoměrně a pro každý faktor musí být součet vah roven 1. Na základě toho syntaxe příkazu ESTIMATE pro obecnou lineární hypotézu odhadu střední hodnoty μ_1 úrovně 1 faktoru A podle rovnice (28) dostane tvar:

```
estimate 'L:mean a1' intercept 1 a 1 0 0 b 0.5 0.5 a*b 0.5 0.5 0 0 0 0;
```

V případě, že obecný lineární model podle (24) se 2 faktory obsahuje také i 1 vysvětlující proměnnou (kovariátu) x_{ijk} , rovnice modelu (24) se změní na tvar:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma x_{ijk} + \varepsilon_{ijk}, \quad (30)$$

kde $i = 1, 2, 3, j = 1, 2$ a $k = 1, \dots, n_{ij}$. Parametr γ odpovídá směrnici regresní přímky.

Vektor odhadovaných parametrů β modelu (30) s 1 kovariátou se změní na:

$$\beta = (\mu \ \alpha_1 \ \alpha_2 \ \alpha_3 \ \beta_1 \ \beta_2 \ (\alpha\beta)_{11} \ (\alpha\beta)_{12} \ (\alpha\beta)_{21} \ (\alpha\beta)_{22} \ (\alpha\beta)_{31} \ (\alpha\beta)_{32} \ \gamma) \quad (31)$$

K odhadu střední hodnoty μ_1 úrovně 1 faktoru A, tj. A_1 , se funkce střední hodnoty vycházející z rovnice modelu se 2 faktory a 1 vysvětlující proměnnou vyjádří ve tvaru:

$$\mu_1 = \mu + \alpha_1 + \frac{\beta_1}{2} + \frac{\beta_2}{2} + \frac{(\alpha\beta)_{11}}{2} + \frac{(\alpha\beta)_{12}}{2} + \bar{x}_{ijk}. \quad (32)$$

Vektor vah L , který odpovídá funkci odhadu podle (32), je definován rovnicí:

$$L = (1 \mid 1 \ 0 \ 0 \mid 0.5 \ 0.5 \mid 0.5 \ 0.5 \ 0 \ 0 \ 0 \ 0 \mid \bar{x}_{ijk}), \quad (33)$$

kde \bar{x}_{ijk} je průměr vysvětlující proměnné x_{ijk} , $i = 1, 2, 3, j = 1, 2$ a $k = 1, \dots, n_{ij}$.

Příkaz CONTRAST má syntax velmi podobnou jako příkaz ESTIMATE. V případě, že jsou váhy vektoru L součástí příkazu CONTRAST, musí jejich součet ve vektoru vah být roven 0. Podobným způsobem je možné odhadovat jakékoliv kombinace lineárních funkcí parametrů lineárního modelu s testováním významnosti takto získaných odhadů pomocí obecných lineárních hypotéz.

4. UKÁZKA ODHADU PARAMETRŮ LINEÁRNÍHO MODELU V SYSTÉMU SAS

Programový systém SAS je vybaven několika desítkami statistických procedur, které jsou předurčeny pro různé typy regresní analýzy. Modely pro regresi mají svůj původ v charakteristikách sledované proměnné (diskrétní nebo spojitá, normálně nebo nenormálně rozdělená), v předpokladech o tvaru modelu (lineární, nelineární nebo zobecněný lineární), o mechanismu generování dat (data pocházející z náhodného výběru, pozorování nebo experimentální data) a v metodě odhadu. Pro analýzu lineárních modelů se převážně využívá procedura REG a procedura GLM. Obě procedury umožňují odhady parametrů lineárních modelů.

Pro ukázkou modelování lineárního modelu a analýzy rozptylu byl zvolen soubor dat CLASS ze systémové knihovny SASHELP. Datový soubor SASHELP.CLASS obsahuje informace o malé fiktivní třídě studentů. Proměnné souboru zahrnují 2 znakové a 4 numerické proměnné a struktura souboru je ilustrována obrázkem č. 1:

Obrázek č. 1: Zobrazení struktury datového souboru SASHELP.CLASS

Variables in Creation Order			
#	Variable	Type	Len
1	Name	Char	8
2	Sex	Char	1
3	Age	Num	8
4	Height	Num	8
5	Weight	Num	8

Zdroj: vlastní zpracování autora podle SAS Institute Inc. 2023b

Pro účely ukázky použití regresní procedury REG bude použit lineární model ve tvaru:

$$Weight_i = Height_i + \varepsilon_i. \quad (34)$$

Modelování obecného lineárního modelu bude realizováno pomocí procedury GLM s využitím singulární parametrizace podle (6). Obecný lineární model tak bude tvořen výrazem:

$$Weight_i = Sex + Height_i + Height_i * Sex + \varepsilon_i. \quad (35)$$

Znaková proměnná Sex bude v modelu (35) sehrávat úlohu klasifikačního faktoru. Tato datová sada se často používá v dokumentaci SAS k ilustraci kódování SAS.

4.1. ODHADY PARAMETRŮ LINEÁRNÍHO MODELU PROCEDUROU REG

Pro odhady parametrů pomocí regresních procedur systému SAS se požadovaný model ve tvaru podle (34) v proceduře REG nastaví příkazem MODEL. Graf modelované závislosti, diagnostické grafy a jiné grafy zajistí příkaz PLOTS. Syntaxe příkazu pro modelování lineární závislosti (pro data CLASS ze systémové knihovny SASHELP) procedurou REG je následující:

```
title 'Simple Linear Regression';
ods graphics on;

proc reg data=sashelp.class plots(unpack);
    model Weight = Height;
run;

quit;
```

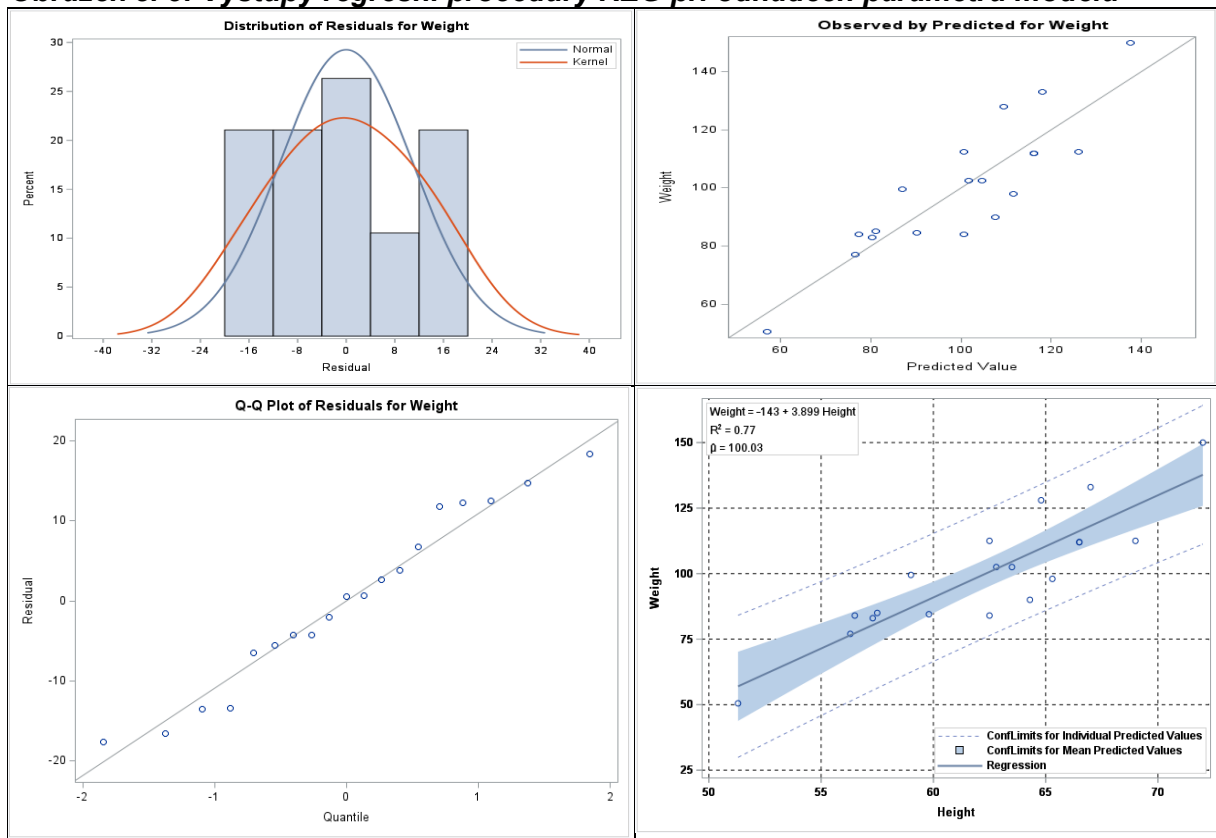
Odhadovaný lineární regresní model (34) obsahuje absolutní člen (intercept) a 1 nezávislou proměnnou HEIGHT. Analýzou rozptylu byla zjištěna hodnota F statistiky v testu významnosti modelu ve výši 57,08, na základě které lze

s pravděpodobností $<0,0001$ zamítnout hypotézu o statistické nevýznamnosti odhadnutého modelu. Odhadnutý regresní model vysvětluje velkou část variability vysvětlované proměnné HEIGHT, o čem vypovídá dosažená hodnota koeficientu determinace R^2 . Zjištěné hodnoty t statistik dovolují zamítnout nulovou hypotézu o nevýznamnosti koeficientu beta u vysvětlující proměnné WEIGHT. Lze zamítnout také i nulovou hypotézu o nevýznamnosti absolutního členu (interceptu) v modelu. Uvedená interpretace statistických testů je oprávněná jen za předpokladu normálně rozdělených chyb v odhadovaném modelu. Informace po provedeném odhadu ze statistické regresní procedury REG jsou ilustrovány na obrázku č. 2.

Obrázek č. 2: Výstupy regresních procedur při odhadech parametrů lineárního modelu

Simple Linear Regression					
Model: MODEL1					
Dependent Variable: Weight					
Number of Observations Read		19			
Number of Observations Used		19			
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	7193.24912	7193.24912	57.08	<.0001
Error	17	2142.48772	126.02869		
Corrected Total	18	9335.73684			
Root MSE		11.22625	R-Square	0.7705	
Dependent Mean		100.02632	Adj R-Sq	0.7570	
Coeff Var		11.22330			
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-143.02692	32.27459	-4.43	0.0004
Height	1	3.89903	0.51609	7.55	<.0001

Zdroj: vlastní zpracování autora

Obrázek č. 3: Výstupy regresní procedury REG při odhadech parametrů modelu

Zdroj: vlastní zpracování autora

Grafické zobrazení závislosti naměřených pozorování a odhadnutého modelu včetně rovnice a vybraných statistik – koeficientu determinace a průměru závisle proměnné je znázorněno na obrázku č. 3. Na obrázku č. 3 jsou vybrány také některé diagnostické grafy, jejichž pomocí se testuje splnění předpokladů, například normalita a nezávislost reziduí.

Použití procedury REG pro analýzu lineárního modelu je jednoduché, intuitivní a uživatelsky dostatečně přívětivé. Statistická procedura REG poskytuje také i velké množství diagnostických nástrojů, které pomáhají v mnoha otázkách regresní analýzy. Podrobnější popis příkazů a parametrů procedury REG a přesné matematické definice jednotlivých dostupných statistik na výstupu regresní procedury REG lze dohledat v originální dokumentaci programového systému SAS [9].

4.2. ODHADY PARAMETRŮ LINEÁRNÍHO MODELU PROCEDUROU GLM

Odhad parametrů obecného lineárního modelu bez klasifikačního faktoru

V případě, že se do obecného lineárního modelu (35) nezahrne klasifikační faktor, lze nahlížet na obecný lineární model jako na lineární model podle vztahu (34). Syntaxe příkazů procedury GLM se stává shodnou se syntaxí příkazů procedury REG. Jediný rozdíl v použití procedury GLM je vložení parametru SOLUTION v příkazu MODEL:

```
title 'Simple Linear Regression';
ods graphics on;
```

```
proc glm data=sashelp.class plots(unpack);
    model Weight = Height / solution;
run;

quit;
ods graphics off;
```

Odhadnuté parametry i ostatní přidružené statistiky mají stejné hodnoty i stejnou interpretaci jako u vykonaného odhadu parametrů procedurou REG.

Odhad parametrů obecného lineárního modelu s klasifikačním faktorem

V případě, že je v modelu klasifikační faktor zahrnut, jedná se o obecný lineární model (35). Proto k odhadování parametrů modelu lze použít výlučně proceduru GLM opět s parametrem SOLUTION v příkazu MODEL:

```
title 'One-Factor ANOVA';
ods graphics on;

proc glm data=sashelp.class plots(unpack);
    class Sex (ref='F');
    model Weight = Sex Height Sex*Height / solution;
run;
ods graphics off;
```

Ve srovnání s příkazy procedury REG je v příkazech procedury GLM příkaz CLASS, kterým se do modelu vkládá klasifikační faktor SEX. Singulární parametrizace v modelu (35) je vyjádřena explicitně výrazem (ref='F'), a proto úroveň F klasifikačního faktoru je úrovní referenční. Jedná se tedy o obecný lineární model v aditivním tvaru s 1 klasifikačním faktorem SEX a jeho interakcí s vysvětlovanou proměnnou HEIGHT.

Výstupem procedury GLM po realizované analýze rozptylu a odhadu parametrů modelu podle (35) jsou výstupy ilustrované na obrázku č. 4. Informace, které procedura GLM ve svém implicitní nastavení představuje, lze rozdělit na několik relativně samostatných částí. Podobně jako u ostatních regresních procedur (i u procedury REG popsané v předešlé podkapitole) jsou nejprve uvedeny informace o počtech pozorování vstupujících do zpracování procedurou GLM. Následují informace o provedené analýze rozptylu vycházející z definičních vztahů tabulky č. 2. Přehled výsledků uzavírá výstup s hodnotami odhadnutých parametrů obecného lineárního modelu. Podobně jako u procedury REG informace na výstupu slouží k testování hypotézy o významnosti parametrů, faktorů i modelu jako celku. Výstupy procedury GLM po analýze obecného lineárního modelu a odhadu jeho parametrů představuje obrázek č. 4.

Obrázek č. 4: Výstupy procedury GLM po analýze rozptylu obecného lineárního modelu

One-Factor ANOVA							
Class Level Information							
Class	Levels	Values					
Sex	2	M F					
Number of Observations Read		19					
Number of Observations Used		19					
Page Break							
One-Factor ANOVA							
Dependent Variable: Weight							
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F		
Model	3	7402.992420	2467.664140	19.15	<.0001		
Error	15	1932.744422	128.849628				
Corrected Total	18	9335.736842					
R-Square Coeff Var Root MSE Weight Mean							
0.792974		11.34821		11.35120		100.0263	
Source	DF	Type I SS	Mean Square	F Value	Pr > F		
Sex	1	1681.122953	1681.122953	13.05	0.0026		
Height	1	5696.840666	5696.840666	44.21	<.0001		
Height*Sex	1	25.028801	25.028801	0.19	0.6657		
Source	DF	Type III SS	Mean Square	F Value	Pr > F		
Sex	1	15.202417	15.202417	0.12	0.7360		
Height	1	5654.260964	5654.260964	43.88	<.0001		
Height*Sex	1	25.028801	25.028801	0.19	0.6657		
Parameter	Estimate		Standard Error	t Value	Pr > t		
Intercept	-117.3697952	B	48.60161448	-2.41	0.0290		
Sex M	-23.7312215	B	69.08843486	-0.34	0.7360		
Sex F	0.0000000	B	.	.	.		
Height	3.4244052	B	0.79971932	4.28	0.0007		
Height*Sex M	0.4881440	B	1.10756565	0.44	0.6657		
Height*Sex F	0.0000000	B	.	.	.		

Poznámka: Výrazy, za jejichž odhady následuje písmeno "B", nejsou jednoznačně odhadnutelné. Matice X'X je singulární, a k řešení normálních rovnic byla použita zobecněná inverze podle [2].

Zdroj: vlastní zpracování autora

Možná interpretace informací z výstupu analýzy obecného lineárního modelu a odhadu parametrů je následující:

V testu významnosti modelu (tzv. nulová hypotéza pro odhadovaný model) byla zjištěna hodnota statistiky F ve výši 19,15, která umožňuje hypotézu o nevýznamnosti odhadovaného modelu zamítnout. Byl zaznamenán vysoký podíl variability vysvětlované proměnné HEIGHT modelem podle (35), čemuž nasvědčuje koeficient determinace na úrovni 0,792974. Při testování hypotézy o významnosti klasifikačního faktoru SEX a interakci tohoto faktoru s numerickou proměnnou WEIGHT nabízí procedura GLM v základních nastaveních k posouzení 2 typů součtů čtverců. Jedná

se o rozklady součtů čtverců, které jsou podle [9] ve výstupech procedury GLM označeny jako Type I SS a Type III SS.

Oba typy rozkladů součtů čtverců vycházejí ze stejných definičních vztahů (podle tabulky č. 2). Rozdíl mezi rozklady sum čtverců typu I a rozklady sum čtverců typu III spočívá ve způsobu zahrnutí faktoru do analyzovaného modelu při testování jeho významnosti. Zatímco u rozkladů podle typu I jsou hodnoty součtu čtverců jednotlivých faktorů závislé na hodnotě součtu čtverců faktoru již zahrnutého v modelu, rozklady součtů čtverců podle typu III na pořadí zahrnutí faktoru do modelu závislé nejsou. V obou typech rozkladů sum čtverců má zdroj variability (faktor nebo interakce), který je do rozkladů typu I vybrán jako poslední, stejnou hodnotu u rozkladů podle typu III. Porovnání součtů čtverců obou typů u jednotlivých faktorů umožňuje lépe kvantifikovat vliv faktorů, resp. jejich interakcí, a tedy lépe posoudit významnost konkrétního faktoru pro daný model⁶.

V tabulce rozkladu čtverců typu I jednotlivé řádky postupně shora dolů udávají, o kolik se přidáním daného členu zmenší reziduální součet čtverců. Obecně tedy závisí na pořadí, v jakém se jednotlivé členy (faktory, interakce) objevují. V každém řádku tedy statistika F (prostřednictvím příslušné dosažené hladiny testu p) vypovídá o významnosti té části variability závisle proměnné, kterou nelze vysvětlit pomocí všech výše uvedených členů a kterou daný faktor či interakce vysvětluje. Rozklad čtverců typu III hodnotí přínos daného faktoru (interakce faktorů) po adjustaci vůči všem ostatním členům bez ohledu na jejich pořadí. Pro tento rozklad je obtížné hledat interpretaci, protože hodnotí vzrůst reziduálního součtu čtverců způsobený vyloučením daného faktoru či interakce, když v modelu zůstanou (je provedena adjustace vůči nim) všechny ostatní faktory či interakce včetně případných interakcí, v nichž je člen (faktor či interakce) obsažen.

Na základě testování významnosti jednotlivých zdrojů variability (faktorů a interakcí) byla u proměnné HEIGHT hypotéza o nevýznamnosti faktoru zamítnuta – a to i podle rozkladů typu I a typu III. U ostatního faktoru, resp. interakce se nedá na základě zjištěných hodnot F statistik nevýznamnost vyloučit. S podobným výsledkem je možné interpretovat i odhady parametrů, kdy se za statisticky významné dají považovat odhadnuté parametry absolutního členu (intercept) a parametru přidruženého k proměnné HEIGHT. Parametrizací modelu z důvodu zahrnutí klasifikační proměnné se matice $X'WX$ stala singulární a k řešení normálních rovnic musela být použita její zobecněná inverzní matice. Podobně jako u procedury REG jsou součástí odhadu procedurou GLM diagnostické grafy pro kontrolu splnění předpokladů lineární regrese.

Porovnání hodnot úrovní klasifikačního faktoru

Důležitým úkolem při analýze dat s klasifikačními efekty je odhadnout typickou hodnotu vysvětlované proměnné pro každou úroveň daného efektu; často je potřebné tyto odhady také porovnat, které úrovně jsou z pohledu vysvětlované proměnné rovnocenné. Procedura GLM se s úkolem porovnávání hodnot na jednotlivých úrovních klasifikačního faktoru vypořádává 2 způsoby, a to:

- pomocí výběrových aritmetických průměrů,
- tzv. průměry nejmenších čtverců (LS -průměry).

⁶ Rozklady čtverců typu I jsou nazvány sekvenčními rozklady součtu čtverců a součty čtverců typu III parciálními.

Tyto statistiky (výběrové aritmetické průměry a *LS*-průměry) obecně nejsou shodné a svým rozdílem odrážejí míru nevyváženosti dat pro analýzu rozptylu, resp. vliv faktorů anebo interakcí. *LS*-průměry jsou podle [3] považovány za odhadované marginální populační střední hodnoty. Porovnání skupin se uskutečňuje na základě příkazu MEANS a LSMENS procedury GLM. Příkladem může být syntaxe:

```
ods graphics on;
proc glm data=sashelp.class plots(unpack)=all;
  class Sex (ref='F') / ;
  model Weight = sex Height sex*height / solution;
  means Sex;
  lsmeans Sex;
  lsmeans Sex / at means;
  lsmeans Sex / at Height=80;
run;

quit;

ods graphics off;
```

Na základě výše uvedené sekvence příkazů procedura GLM příkazem MEANS počítá průměry pro všechny úrovně faktoru SEX. Příkaz LSMEANS odhaduje střední hodnoty, a to marginální pro úrovně faktoru SEX. Obsahuje-li příkaz LSMEANS také parametr AT, procedura GLM provádí odhady v příslušné hodnotě, resp. úrovni faktoru, např. příkaz LSMEANS Sex / AT HEIGHT = 80 odhaduje marginální průměr pro HEIGHT = 80.

5. ZÁVĚR

Cílem článku bylo přiblížit význam a možnosti programového systému SAS z pohledu statistického modelování. Statistické modelování prošlo v posledních desetiletích bouřlivým vývojem, který byl urychlován neustálým rozvojem moderních technologií a rozšiřováním neustále se zkvalitňující výpočetní techniky. Z celé široké oblasti modelování ve statistice byl ve článku položen důraz na klasické a obecné lineární modely, protože patří v oblasti statistického modelování k těm nejdůležitějším.

Článek zavádí nezbytné základní pojmy a definice, a tak poskytuje teoretický úvod do sledované tematiky. Na několika vybraných příkladech byly demonstrovány ukázky modelovací techniky procedurami REG a GLM, které tvoří pilíře regresních technik programového systému SAS. Výstupy z regresní analýzy obou procedur byly dostatečně interpretovány tak, aby poskytly všem výzkumníkům a datovým analytikům návod k dalšímu zkoumání.

LITERATURA

- [1] BELSLEY, D.A. – KUH, E. – WELSCH, R. E.: Regression diagnostics: Identifying Influential Data and Sources of Collinearity.. New York: John Wiley and Sons. 1980. 321 s. ISBN ISBN 0-47 1-69 1 17-8.
- [2] BEN-ISRAEL, A. – GREVILLE, T. N. E.: Generalized inverses: theory and applications. Springer-Verlag New York, 2003. Inc.:New York. 436.s. ISBN 0-387-00293-6.

- [3] LITTELL, C. L. – STROUP, W. W. – FREUND, R. J.: SAS for Linear Models, 2002. 4th ed. Cary, NC: SAS Institute Inc. xv + 478 s. ISBN 978-1-59047-023-7.
- [4] LITTELL, R. C.: The Evolution of Linear Models in SAS: A Personal Perspective. SAS Global Forum 2011. [cit. 02. 07. 2023]. Dostupné na: <http://support.sas.com/resources/papers/proceedings11/325-2011.pdf>
- [5] McCULLAGH, P. – NELDER, J. A.: Generalized Linear Models. 2nd ed. London: Chapman & Hall, 1989. 526 s. ISBN 0-412-31760-5.
- [6] NELDER, J.A. – WEDDERBURN, R.W.M.: Generalized Linear Models. In: Journal of the Royal Statistical Society. Series A (General). Vol. 135, No. 3 1972, pp. 370 – 384.
- [7] RUTHERFORD, A.: ANOVA and ANCOVA: a GLM approach, 2nd Edition. John Wiley & Sons, Inc.: Hoboken, New Jersey. 2011. xvi + 344 s. ISBN 978-0-470-38555-5.
- [8] SAS Institute Inc. 2023a. SAS/STAT® 15.3 User's Guide. Chapter 1 Procedures by Category. Cary, NC: SAS Institute Inc. [cit. 02. 07. 2023]. Dostupné na: <http://support.sas.com/documentation/onlinedoc/stat/153/statug.pdf>.
- [9] SAS Institute Inc. 2023b. SAS/STAT® 15.3 User's Guide. Chapter 4 Introduction to Statistical Modelling with SAS/STAT Software. Cary, NC: SAS Institute Inc. s. 26. [cit. 03. 07. 2023].
Dostupné na: <http://support.sas.com/documentation/onlinedoc/stat/153/statug.pdf>.
- [10] TOBIAS, R. – KIERNAN, K. – TAO, J. – GIBBS, P.: CONTRAST and ESTIMATE Statements Made Easy: The LSMESTIMATE Statement. SAS Global Forum 2011. [cit. 22. 08. 2023]. Dostupné na: <https://support.sas.com/resources/papers/proceedings11/351-2011.pdf>.

RESUMÉ

Význam statistického modelování v analýze dat se v posledních několika desetiletích rapidně zvýšil. Při zkoumání přírodních, či ekonomických nebo i jiných dat již nestačí pouhý popis těchto dat nebo klasická statistická inference o datech. V současnosti se do popředí zájmu datových analýz dostávají takové metody sofistikovaného zpracování dat, které umožňují pochopit mechanismus vzniku sledovaných dat, hledat a nacházet v datech nové poznatky a souvislosti mezi nimi a s rozumnou pravděpodobností dovolují předvídat vývoj pozorovaného jevu do určitého předem stanoveného horizontu. V průběhu let se navíc změnila i samotná sledovaná data. V důsledku rychlého vývoje výpočetní techniky a obecné digitalizace života se objem zkoumaných dat neustále zvyšuje, data jsou rozmanitější a jsou získávána vyšší rychlostí. Ze zkoumaných dat se tak zpravidla stávají větší a komplexnější soubory údajů, které pocházejí z nových zdrojů.

Prostředí, které je schopné čelit výše uvedeným výzvám v této oblasti, je právě programový systém SAS. SAS je analytický modulární systém, který již ve svém základním sestavení nabízí svým uživatelům velmi silné a výkonné funkcionality k analýzám a zpracování zkoumaných dat. Například pro statistické modelování tento analytický systém zahrnuje několik desítek statistických procedur, přičemž každá z nich se specializuje pro určité třídy statistického modelování. Příkladem takových funkcionalit, ve kterých jsou implementovány nejdůležitější poznatky moderní regresní vědy, jsou procedury REG a GLM.

RESUME

The importance of statistical modelling in data analysis has increased rapidly over the last few decades. When examining natural or economic or even other data, it is no

longer sufficient to simply describe the data or to make classical statistical inference about the data. Nowadays, data analysis is becoming more interested in sophisticated data processing methods that allow to understand the mechanism of the observed data, to search for and find new insights and connections between data and to predict with reasonable probability the development of the observed phenomenon within a certain predetermined horizon. In addition, the observed data itself has changed over the years. As a result of the rapid development of computer technology and the general digitisation of life, the volume of data examined is constantly increasing, the data are more varied and are being acquired faster. As a result, the examined data are generally larger and more complex data sets, arising from new sources.

An environment that is capable of meeting the above challenges in this area is the SAS software system. SAS is an analytical modular system that, already in its basic setup, offering its users very powerful functionalities for analysing and processing the data under study. For example, for statistical modelling, this analytical system includes several dozens of statistical procedures, each specialized for a particular class of statistical modelling. The examples of such functionalities, which implement the most important findings of the modern regression science, are the REG and the GLM procedures.

PROFESIJNÝ ŽIVOTOPIS

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

KONTAKT

roman.pavelka@statistics.sk

Informácia/Information

TRETIE KOLO PEER REVIEW 19. – 23. jún 2023

THIRD ROUND OF PEER REVIEW 19. – 23. June 2023

V dňoch 19. až 23. júna sa v Štatistickom úrade SR uskutočnila návšteva tímu expertov v rámci Peer review, tzv. partnerského preskúmania štatistických úradov a národných štatistických systémov členských štátov Európskej únie a krajín EZVO (Európske združenie voľného obchodu).



Išlo už o tretie kolo partnerského preskúmania: prvé kolo prebehlo ešte v rokoch 2006 až 2008, druhé kolo v rokoch 2013 až 2015 a aktuálne tretie kolo v rokoch 2021 až 2023.

Prvotným cieľom partnerského preskúmania je posilniť integritu, profesionálnu nezávislosť a zodpovednosť štatistických orgánov Európskeho štatistického systému (EŠS). Ako reakcia na rastúci trend dopytu po okamžitých informáciách a nových výzvach spôsobených mimoriadnymi okolnosťami a neustálou potrebou rýchlejších a zároveň kvalitnejších údajov sa európska oficiálna štatistika prioritne zameriava na kvalitu. Základom spoločného rámca kvality EŠS je Kódex postupov pre európsku štatistiku (ES CoP), predstavujúci súbor noriem kvality, ktoré pôsobia ako samoregulačný nástroj prostredníctvom 16 zásad rozdelených do oblastí inštitucionálneho prostredia, štatistických procesov a výstupov štatistík. Peer review ako jeden z nástrojov zabezpečenia kvality európskej štatistiky kontroluje súlad národných štatistických úradov a iných vnútroštátnych orgánov zodpovedných za rozvoj, tvorbu a šírenie oficiálnej štatistiky so zásadami Kódexu postupov.

Celý proces preskúmania bol rozdelený do štyroch fáz. V prvej fáze vyplnil Štatistický úrad SR tzv. sebahodnotiaci dotazník (SAQ), ktorý následne experti zanalyzovali a vyhodnotili. Spolu s dotazníkom bolo zaslané množstvo relevantných dokumentov prevažne v anglickom jazyku priamo tímu expertov, ktorí boli vybraní na preskúmanie Štatistického úradu SR, ako aj pracovnej skupine pre partnerské preskúmanie a kvalitu v rámci Eurostatu. Nasledovala návšteva expertov, ktorej výsledkom je záverečná správa s odporúčaniami na zlepšenie pre danú krajinu. Národný štatistický úrad má za úlohu vypracovať konkrétne opatrenia s cieľom naplniť odporúčania, ktoré sa budú každoročne monitorovať.

Návšteva tímu expertov 19. – 23. jún 2023

Prvý deň návštevy bol zameraný na predstavenie národného štatistického systému, jeho stavu, zloženia a legislatívnemu základu. Nato bola expertom prezentovaná koordinácia a spolupráca Štatistického úradu SR s jednotlivými členmi národného štatistického systému (NŠS).

Po tomto úvodnom oboznámení sa s NŠS nasledovali jednotlivé sedenia zamerané vždy na konkrétne zásady Kódexu postupov a ich dodržiavania Štatistickým úradom

SR. Prvý blok bol venovaný odbornej nezávislosti a zákonu č. 540/2001 Z. z. o štátnej štatistike a súvisiacej legislatíve z pohľadu zásad odbornej nezávislosti, oprávneniu na zber údajov, štatistickej dôvernosti a ochrane údajov, ako aj nestrannosti a objektívnosti.

Druhý blok bol zameraný na plánovanie a zdroje z pohľadu zásad odbornej nezávislosti, primeranosti zdrojov, zamedzenia nadmernej záťaže respondentov a efektívnosti vynaložených nákladov.

V druhý deň návštevy sa experti v úvodnom bloku zaoberali kvalitou, konkrétnejšie nástrojmi kvality a monitorovaním kvality. Prioritne sa ich otázky zamerali na súlad so zásadami týkajúcimi sa štatistických výstupov, t. j. záväzkov ku kvalite, relevantnosť, presnosť a spoľahlivosť, včasnosť a časová presnosť, koherentnosť a porovnateľnosť a prístupnosť a zrozumiteľnosť.

V druhom bloku bola diskusia venovaná diseminácii a štatistickej dôvernosti s ohľadom na štatistickú dôvernosť a ochranu údajov, nestrannosť a objektívnosť a prístupnosť a zrozumiteľnosť.

Hlavnou témou tretieho bloku bola metodológia, zber údajov, spracovanie údajov a administratívne údaje z pohľadu zásad oprávnenia na zber údajov, správna metodika a vhodné štatistické postupy.

Súčasťou návštevy bol okrem jednotlivých blokov, na ktorých sa zväčša – nie však výlučne – zúčastňuje prevažne vrcholový a stredný manažment, aj blok venovaný juniorským zamestnancom, t. j. zamestnancom, ktorý v Štatistickom úrade SR pôsobia 2 – 5 rokov. Stretnutie prebehlo v maximálne priateľskom a profesionálnom duchu, ostatne tak, ako aj všetky ostatné stretnutia. Expertov zaujímal pohľad profesijne a v niektorých prípadoch aj vekovo mladších zamestnancov na súlad práce Štatistického úradu SR s jednotlivými zásadami Kódexu postupov a status úradu ako poskytovateľa oficiálnej štatistiky v spoločnosti.

Záver druhého dňa patril progresu úradu za posledné tri roky a plánom na nasledujúcich päť rokov, na ktorom vrcholový manažment úradu oboznámil expertov Eurostatu so smerovaním úradu z pohľadu inštitucionálneho prostredia, štatistických procesov a štatistických výstupov.

Tretí deň bol celý venovaný vybraným členom národného štatistického systému, ktorí rovnako ako Štatistický úrad SR vopred zaslali vyplnený sebahodnotiaci dotazník za svoju inštitúciu a boli pripravení predstaviť expertom svoju účasť na rozvoji, tvorbe a šírení oficiálnej štatistiky a súlad s Kódexom postupov pre európsku štatistiku. Na základe špecifických kritérií (napr. dôležitosť pre európsku štatistiku) boli zvolené tri iné vnútroštátne orgány, ktoré sú členmi NŠS – menovite Slovenský hydrometeorologický ústav, Migračný úrad Ministerstva vnútra SR a Správa štátnych hmotných rezerv SR, ktorým boli venované jednotlivé bloky tretieho dňa. Každá zo spomenutých inštitúcií okrem súladu s Kódexom postupov hodnotila aj úlohu Štatistického úradu SR ako koordinátora NŠS.

Záver tretieho dňa zaplnil blok venovaný Koordinačnej rade pre štátnu štatistiku ako jednému z nástrojov koordinácie, ktorá bola zriadená pre potreby týkajúce sa plnenia úloh štátnej štatistiky inými orgánmi vykonávajúcimi štátnu štatistiku.

Štvrtý deň boli stretnutia zamerané na partnerov Štatistického úradu SR. Experti sa stretli so zástupcami hlavných poskytovateľov administratívnych zdrojov údajov (Ministerstvo financií SR, Národné centrum zdravotníckych informácií); poskytovateľov nových zdrojov údajov (Národná diaľničná spoločnosť, COOP Jednota, O2); ako aj so zástupcami hlavných používateľov štatistických údajov – médií (Zdravotnícke noviny, Hospodárske noviny, Retail magazín); ministerstiev a ďalších verejných i súkromných inštitúcií (Národná banka Slovenska, Ministerstvo financií SR, Ministerstvo dopravy SR, Inštitút pre výskum práce a rodiny); obchodných asociácií a odborových zväzov (Asociácia hotelov a reštaurácií Slovenska, Zväz stavebných podnikateľov Slovenska, Zväz automobilového priemyslu) a vedeckej komunity (Slovenská akadémia vied, Fakulta hospodárskej informatiky Ekonomickej univerzity v Bratislave, Infostat).

Posledný deň návštevy sa uskutočnilo stretnutie expertov s vrcholovým manažmentom Štatistického úradu SR a zástupcami zúčastnených iných vnútroštátnych orgánov, na ktorom experti zhrnuli svoje závery a predstavili prvý návrh odporúčaní.

Proces tretieho kola Peer review ešte nie je ukončený. Po oficiálnom vydaní správy na stránke Eurostatu v novembri tohto roka bude nasledovať vypracovanie opatrení reflektujúcich odporúčania expertov, ktoré sa budú v priebehu času preskúmať.

Na záver si dovoľíme tvrdiť, že celý priebeh partnerského preskúmania od prvého odoslaného dokumentu po návštevu tímu expertov na pôde Štatistického úradu SR bol maximálne profesionálne zvládnutý, čo potvrdili aj všetci zúčastnení.

Partnerské preskúmanie takéhoto charakteru je okrem iného aj vynikajúcou príležitosťou pre zamestnancov ŠÚ SR na všetkých úrovniach získať neoceniteľné skúsenosti a prax. Najbližšie kolo Peer review by sa malo konať o ďalších 7 – 8 rokov.



Zdroj: Štatistický úrad SR

Ing. Helena GLASER-OPITZOVÁ
Bc. Albert IVANČÍK

Autori pracujú v sekcii všeobecnej metodiky, registrov a koordinácie národného štatistického systému Štatistického úradu SR.

Informácia/Information

AMSE 2023 – 25. ROČNÍK KONFERENCIE

AMSE 2023 – 25TH YEAR OF THE CONFERENCE

Konferencia **AMSE 2023 (Applications of Mathematics and Statistics in Economics 2023)**, ktorá sa konala v Rajeckých Tepliciach na sklonku augusta a začiatku septembra tohto roku, oslávila už jubilejné dvadsiate piate výročie. Konferenciu AMSE organizujú na rotačnej báze akademické pracoviská z troch susediacich krajín: Ekonomická fakulta Univerzity Mateja Bela v Banskej Bystrici (za Slovensko), súčasná Fakulta manažmentu, počítačových vied a financií Ekonomickej univerzity vo Vroclave (za Poľsko) a Fakulta informatiky a štatistiky Vysokiej školy ekonomickej v Prahe (za Českú republiku). Konferencia pozostávala z dvoch častí. Formálna časť venovaná prezentácii výsledkov vedeckej práce a akademickej diskusii sa konala vo štvrtok 31. augusta 2023 a v piatok doobeda 1. septembra 2023. Neformálna časť zameraná na kultúrne a turistické poznávanie blízkeho prostredia Strážovských vrchov pokračovala v piatok poobede 1. septembra 2023 a v sobotu 2. septembra 2023.

Vedecký priebeh konferencie garantoval vedecký výbor koncipovaný akademickými osobnosťami zo všetkých troch zaangažovaných pracovísk, ktorý pracoval pod vedením Rudolfa Zimku. Organizáciu konferencie zaštili pracovníci Katedry kvantitatívnych metód a informačných systémov Ekonomickej fakulty Univerzity Mateja Bela v Banskej Bystrici pod vedením Petra Laca.

Konferencia mala skromný začiatok v roku 1998, keď v Liptovskom Trnenci Miroslav Abrahám z Ekonomickej fakulty Univerzity Mateja Bela v Banskej Bystrici navrhol zorganizovať stretnutie učiteľov matematiky a štatistiky, ktorí cítili potrebu vylepšiť svoj spôsob výkladu kvantitatívnych metód pri výučbe ekonómie. Už prvé stretnutie v Liptovskom Trnenci malo medzinárodný charakter a bolo frekventované kolegami z Vysokiej školy ekonomickej v Prahe a účastníkom z Univerzity v Groningene v Holandsku. V priebehu nasledujúcich rokov počet účastníkov pravidelne rástol a od štvrtého ročníka sa ujala rotačná organizácia medzi troma kooperujúcimi pracoviskami.

Počas 26 rokov svojho trvania sa konferencia AMSE konala celkovo 25-krát a jej kontinuita bola krátkodobo narušená v dôsledku covidových udalostí v roku 2020. V nasledujúcom roku 2021 sa konferencia uskutočnila v online formáte. Keďže v prvých troch rokoch sa podujatie opakovane uskutočnilo na Slovensku, celkovo sa konalo 10 ročníkov na Slovensku (Liptovský Trnovec dvakrát, Poprad dvakrát, Banská Bystrica, Demänovská Dolina, Gerlachov, Banská Štiavnica, Nižná, Rajecké Teplice), 8 ročníkov v Českej republike (Zadov, České Budějovice, Trutnov, Uherské Hradiště, Liberec, Jindřichův Hradec, Kutná Hora, Velké Losiny) a 7 ročníkov v Poľsku (Kudowa-Zdrój, Wrocław, Wisła, Łądek-Zdrój, Jerzmanowice, Szklarska Poręba, Wrocław online). Konferencia sa začala pod názvom Applications of Mathematics and Statistics in Economy (z anglického pojmu political economy, teda politická ekonómia) a od roku 2014 sa názov pozmenil podľa súčasného jazykového úzu na Applications of Mathematics and Statistics in Economics. Skratka AMSE, pod ktorou je konferencia známa, však zotrváva celý čas. Po odchode zakladateľa konferencie sa ako vedúce

osobnosti za slovenskú stranu vyprofilovali Rudolf Zimka a Emília Zimková. Za českú stranu sú nestormi konferencie Richard Hindls a Stanislava Hronová a za poľskú stranu Walenty Ostasiewicz a Joanna Dębicka. Na všetkých 25 ročníkoch sa zúčastnili Rudolf Zimka, Richard Hindls, Stanislava Hronová a Peter Laco, ktorí boli na konferencii dekorovaní slávnostným ocenením a vyjadrením vďaky. Na podujatí za ostatné obdobie štandardne vystupujú so svojimi plenárnymi prednáškami predstavitelia Českého štatistického úradu a Štatistického úradu Slovenskej republiky.

Na dvadsiatom piatom ročníku sa prezentovalo 41 účastníkov a popri dvoch plenárnych prednáškach odznelo 25 vedeckých príspevkov od 53 rôznych autorov a spoluautorov zo Slovenska, Českej republiky a Poľska. Konferenciu svojou účasťou podporil predseda Českého štatistického úradu Marek Rojíček, generálne riaditeľky Štatistického úradu Slovenskej republiky Ľudmila Ivančíková a Helena Glaser-Opitzová a predsedníčka Slovenskej štatistickej a demografickej spoločnosti Iveta Stankovičová.



Zdroj: vlastné fotografie

Konferenciu otvoril 31. augusta 2023 predseda vedeckého výboru Rudolf Zimka a Emília Zimková v zastúpení rektora Univerzity Mateja Bela v Banskej Bystrici Vladimíra Hladovského predniesla jeho otvárací príhovor. Nasledovala plenárna sekcia, na ktorej vystúpil s prednáškou predseda Českého štatistického úradu Marek Rojíček. Po ňom generálna riaditeľka Štatistického úradu Slovenskej republiky Ľudmila Ivančíková predniesla prednášku, ktorú pripravila s predsedom Štatistického úradu Slovenskej republiky Petrom Peťkom. V ten deň sa uskutočnili ešte štyri sekcie venované postupne (1.) štúdiu efektívnosti a ekonomickým analýzám, (2.) ekonomickým otázkam a oficiálnej štatistike, (3.) ekonomickej demografii a demografickej ekonomii a (4.) štatistickým modelom a technikám.

Nasledujúci deň 1. septembra 2023 sa uskutočnili iba dve sekcie zamerané na (1.) aplikácie štatistiky a penzijné systémy a (2.) dynamické modely a históriu štatistiky. Konferenciu uzavrel predseda vedeckého výboru Rudolf Zimka spolu s Richardom

Hindlsom za českú stranu a Joannou Dębickou za poľskú stranu. Záver konferencie sa tradične spája s avízom miesta a času konania budúročného podujatia a tento symbol kontinuity bol dodržaný aj teraz. Konferenciu AMSE 2024 bude organizovať Fakulta manažmentu, počítačových vied a financií Ekonomickej univerzity vo Vroclave. Joanna Dębicka a Beata Zmysłona v mene poľskej časti vedeckého výboru adresovali všetkým účastníkom pozvanie do Pawłowic pri Vroclave, kde sa bude konať v dňoch 28. augusta 2024 až 1. septembra 2024 už dvadsiaty šiesty ročník.

V druhej časti konferenčného dňa účastníci podnikli vychádzku na Lietavský hrad a posledný deň konferencie 2. septembra 2023 mali na výber tri výlety rôznej náročnosti, zahrňujúce prehliadku Sobášneho paláca v Bytči a turistické aktivity v Súľovských skalách.

Konferencia AMSE 2023 preukázala svoju zrelosť. Stretli sa na nej príslušníci viacerých generácií vo vekovom rozpätí od zakladajúcich členov až po mladých tvorivých pracovníkov vstupujúcich do akademického sveta. Význam konferencie pre akademickú sféru dokázal živý záujem participantov z viacerých vysokoškolských a vedeckých pracovísk na Slovensku, v Českej republike a Poľsku. Vzťah preberaných tém k oficiálnej štatistike dosvedčili svojou účasťou predstavitelia národných štatistických úradov v Českej republike a na Slovensku. Konferencia vytvorila platformu pre disemináciu vedeckých poznatkov a poskytla účastníkom námety na riešenie teoretických a praktických problémov ekonómie rigoróznymi metódami matematiky a štatistiky.

doc. PhDr. Ing. Martin BOĎA, PhD.
prof. RNDr. Rudolf ZIMKA, CSc.

Autori sú akademickými pracovníkmi na Katedre kvantitatívnych metód a informačných systémov Ekonomickej fakulty Univerzity Mateja Bela v Banskej Bystrici.

PRIPRAVUJEME/COMING SOON

***Monotematické číslo Slovenskej štatistiky a demografie 1/2024 na tému
Socioekonomické aspekty BigData.
Monothematic issue of the Slovak Statistics and Demography 1/2024 on the topic
Socioeconomic aspects of BigData.***

* * *

***ONLINE VERZIA ČÍSLA 4/2023 SLOVENSKEJ ŠTATISTIKY A DEMOGRAFIE JE
VEREJNE DOSTUPNÁ na internetovej stránke ssad.statistics.sk od 15. OKTÓBRA
2023.***

***THE ONLINE VERSION OF THE JOURNAL SLOVAK STATISTICS AND
DEMOGRAPHY No 4 (2023) IS PUBLICLY BE AVAILABLE at the website
ssad.statistics.sk from OCTOBER 15, 2023.***

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.

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

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

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

INFORMATION FOR AUTHORS

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

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

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

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

Articles are reviewed. The opponent procedure is mutually anonymous. The final decision on the article's publication is made by the editorial board. The editorial office reserves the right to publish articles approved by the editorial board in printed and electronic form at the website ssad.statistics.sk.

SLOVENSKÁ ŠTATISTIKA A DEMOGRAFIA

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

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

Vydáva:

Štatistický úrad SR

Identifikačné číslo vydavateľa:

IČO 00166197

Vychádza:

Štyrikrát ročne

Dátum vydania:

15. október 2023

Tlač:

Reprografické stredisko
Štatistického úradu SR

Predplatné:

20 € (na rok)
5 € (za jeden výtlačok)

Objednávky prijíma:

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

SLOVAK STATISTICS AND DEMOGRAPHY

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

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

Issued by:

Statistical Office of the SR

Company registration number:

00166197

Published:

Four times a year

Date of issue:

15th October 2023

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