

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

3/2026

ročník/volume 36

Recenzovaný vedecký časopis so zameraním na prezentáciu moderných štatistických a demografických metód a postupov.

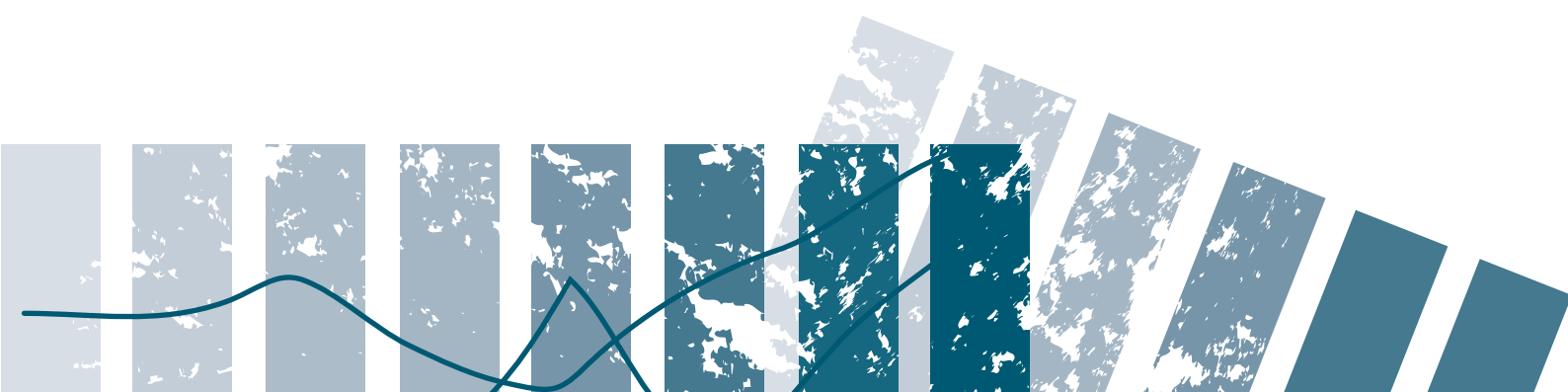
Scientific peer-reviewed journal focusing on the presentation of modern statistical and demographic methods and procedures.

Článok/Article: 2

Typ článku/Type of article: vedecký článok/scientific article

Strany/Pages: 21 – 36

Dátum vydania/Publication date: 15. júl 2026/July 15, 2026



Boris FRANKOVIČ, Monika RÁKOŠOVÁ
Štatistický úrad Slovenskej republiky

GEO-REFERENCOVANÁ VÝBEROVÁ VZORKA ZISŤOVANIA SILC MRK

GEO-REFERENCED SAMPLE OF THE SILC MRK SURVEY

ABSTRAKT

Údaje členené podľa národnosti predstavujú nevyhnutný predpoklad na tvorbu verejných politík založených na dátach. Napriek tomu, že mnohé politiky sú zamerané na marginalizované komunity, dostupnosť a kvalita údajov o rómskej populácii zostáva výzvou. Štatistický úrad Slovenskej republiky kontinuálne realizuje štatistické zisťovanie o príjmoch a životných podmienkach domácností (EU SILC). Toto zisťovanie sa zameriava na celkovú referenčnú populáciu súkromných domácností bez osobitného rozlíšenia alebo zamerania na konkrétne národnostné skupiny. Na základe analogickej metodiky sa realizuje špeciálne zisťovanie SILC MRK, ktoré slúži na kvantifikáciu chudoby a sociálneho vylúčenia v prostredí marginalizovaných rómskych komunít. Cieľovú populáciu predstavujú všetky obývané domácnosti kategorizované podľa jednotlivých typov osídlení. Relevantným zdrojom údajov na vytvorenie opory výberu je databáza obývaných bytov zo Sčítania obyvateľov, domov a bytov 2021 (SODB). Vzhľadom na to, že táto údajová základňa neobsahuje informácie o type osídlenia definovanom v Atlase rómskych komunít, je nevyhnutné prostredníctvom geopriestorovej zhľukovej analýzy klasifikovať jednotlivé domácnosti SODB do príslušných lokalít podľa typu umiestnenia. Takto vytvorená databáza umožňuje realizáciu pravdepodobnostného náhodného výberu, ktorý je nevyhnutným predpokladom na zabezpečenie vysokej kvality výsledkov zisťovania SILC MRK.

ABSTRACT

Data disaggregated by ethnicity constitute a necessary prerequisite for evidence-based policy making. Despite the fact that many public policies are targeted at marginalised communities, the availability and quality of data on the Roma population remain a challenge. The Statistical Office of the Slovak Republic continuously conducts the Survey on income and living conditions (EU-SILC). This survey focuses on the overall reference population of private households without any specific distinction or focus on particular ethnic groups. Based on an analogous methodology, the specialised SILC MRK survey is conducted to measure poverty and social exclusion in marginalised Roma communities. The target population consists of all occupied households categorised according to individual settlement types. A relevant data source for constructing the sampling frame is the Database of occupied dwellings from the 2021 Population and Housing Census. Since this data source does not contain information on settlement types defined in the Atlas of Roma Communities, it is necessary to classify individual Census households into relevant localities according to settlement type using geospatial cluster analysis. The resulting database enables the implementation of probability random sampling, which is an essential prerequisite for ensuring the high quality of the SILC MRC survey results.

KLÚČOVÉ SLOVÁ

marginalizované rómske komunity, sčítanie obyvateľov, výberová vzorka, priestorová zhľuková analýza

KEY WORDS

marginalized Roma communities, census, sample, geospatial cluster analysis

1. ÚVOD

Cieľom tohto príspevku je analyzovať špecifické aspekty výberu vzorky a zberu údajov medzi osobitnými skupinami populácie, ktoré je ťažké osloviť štandardnými metódami, a to najmä z pohľadu metodického zabezpečenia reprezentatívnosti a vysokej kvality výsledkov príslušného štatistického zisťovania.

Jedným z významných demografických a sociálnych špecifik Slovenskej republiky je relatívne vysoký podiel obyvateľov rómskej národnosti. Preto je pri formovaní rozvojových stratégií a politík nevyhnutné venovať osobitnú pozornosť aj tejto skupine obyvateľstva a podporovať jej sociálne začlenenie a zlepšovanie životných podmienok. Ako uvádza Stratégia rovnosti, inklúzie a participácie Rómov do roku 2030, rómska populácia na Slovensku je veľmi rôznorodá, a preto by bolo stigmatizujúce považovať všetkých Rómov a Rómky za znevýhodnených z hľadiska chudoby a sociálneho vylúčenia. Rovnako nesprávny je tento predpoklad aj vo vzťahu k obyvateľom všetkých osídlení považovaných za rómske. Zároveň však platí, že Rómovia patria medzi tie skupiny obyvateľstva, ktoré sú najviac zraniteľné, a štrukturálne znevýhodnenia a diskriminácia sa ich môžu týkať bez ohľadu na ich socioekonomické zázemie či životné podmienky.

Prioritou Stratégie rovnosti, inklúzie a participácie Rómov do roku 2030 je dosiahnutie viditeľnej zmeny v oblasti rovnosti a začleňovania Rómov. Najdôležitejšími oblasťami sú zamestnanosť, vzdelávanie, zdravie a bývanie. Na prijatú stratégiu nadväzujú viaceré akčné plány, ktoré obsahujú konkrétne opatrenia a aktivity z jednotlivých tematických oblastí. Kľúčovým predpokladom úspešnej realizácie a priebežného odpočtovania plnenia akčných plánov je existencia relevantných a pravidelne aktualizovaných údajov, bez ktorých nie je možné spoľahlivo monitorovať vývoj situácie ani vyhodnocovať dosah prijatých opatrení (Úrad splnomocnenca vlády SR pre rómske komunity, 2021). Jednou z aktivít je aj realizácia Výberového zisťovania o príjmoch a životných podmienkach obyvateľstva rómskych marginalizovaných komunít (SILC MRK).

Štatistický úrad Slovenskej republiky (ŠÚ SR) v zmysle nariadenia¹ realizuje každoročne štatistické zisťovanie o príjmoch a životných podmienkach domácností (EU SILC). Ide o harmonizované zisťovanie, ktorého hlavným cieľom je poskytovanie pravidelných podrobných informácií o príjmoch, úrovni a štruktúre chudoby a sociálnom vylúčení domácností na Slovensku. Výstupy zo zisťovania sú dôležitým podkladom v rozhodovacích procesoch, na koncepčné zámery, analyzovanie sociálnej situácie v domácnostiach a následné prijímanie konkrétnych opatrení smerujúcich k zvyšovaniu kvality života obyvateľov SR. Opora výberu zisťovania EU SILC je založená na databáze obývaných súkromných domácností vytvorenej na základe databázy bytov zo Sčítania obyvateľov, domov a bytov 2021 (SODB). Inštitucionálne, nekonvenčné, neobývané, ako aj niektoré ťažko dostupné domácnosti sú z metodických dôvodov vylúčené. Tento prístup však môže znižovať mieru pokrytia obyvateľov marginalizovaných rómskych komunít (MRK), a to najmä tých, ktoré

¹ Nariadenie Európskeho parlamentu a Rady (EÚ) 2019/1700 z 10. októbra 2019, ktorým sa zavádza spoločný rámec pre európske štatistiky o osobách a domácnostiach na základe individuálnych údajov zbieraných zo vzoriek.

žijú v segregovaných osadách. Z toho dôvodu bolo potrebné hľadať alternatívne riešenie v podobe realizácie samostatného štatistického zisťovania s prispôbenou metodikou.

Výberové zisťovanie SILC MRK sa realizuje od roku 2018 v spolupráci s ÚSVRK. Jeho cieľom je monitorovať príjmy a životné podmienky osôb žijúcich v marginalizovaných rómskych komunitách. Metodologickým základom zisťovania je zisťovanie EU SILC doplnené o otázky špecifické pre MRK a modul diskriminácie vypracovaný v spolupráci s Agentúrou Európskej únie pre základné práva. Jedným z cieľov je, aby boli údaje a výsledky z tohto zisťovania plne porovnateľné so štandardným zisťovaním EU SILC a aby mohli byť integrované do európskeho rámca monitorovania politik integrácie rómskej populácie. Ako uvádza Škobla (2025), zisťovanie umožňuje systematicky kvantifikovať rozsah chudoby, sociálneho vylúčenia a nerovností v prostredí MRK a zároveň tieto ukazovatele porovnávať s bežnou populáciou. To prináša doteraz neexistujúcu možnosť presného merania rozdielov v kľúčových socioekonomických ukazovateľoch, ako sú príjem, zamestnanosť, prístup k bývaniu, zdravotná starostlivosť, vzdelanie či sociálne transfery.

V roku 2024 sa začala metodická príprava tretej vlny zisťovania, v ktorej sa realizoval nový prístup ku konštrukcii opory výberu. Cieľom bolo vytvoriť presný obraz referenčnej populácie domácností MRK podľa jednotlivých kategórií typov osídlení definovaných v Atlase rómskych komunit, ktorý by umožnil vygenerovanie reprezentatívnej vzorky metódou pravdepodobnostného náhodného výberu. Výsledky zisťovania tak poskytnú kvalitné a presné odhady, ktoré sú kľúčovými vstupmi na implementáciu akčného plánu Stratégie rovnosti, inklúzie a participácie Rómov do roku 2030.

2. STRATÉGIA NA ZVÝŠENIE MIERY ODPOVEDÍ

Zber údajov zisťovania SILC MRK sa realizuje najmä prostredníctvom metódy PAPI. Tento spôsob získavania údajov bol zvolený predovšetkým z dôvodu životných podmienok marginalizovaných rómskych komunit, kde je prístup k informačným a komunikačným technológiám obmedzený. Sieť opytovateľov pozostáva z externých zamestnancov – prevažne sociálnych pracovníkov – a interných supervízorov pracujúcich na základe zmlúv. Sociálni pracovníci zohrávajú kľúčovú úlohu pri zlepšovaní životných podmienok a sociálnej inklúzie MRK. Vzhľadom na špecifické bariéry, ktorým tieto komunity čelia, ako sú nízka úroveň vzdelania, nezamestnanosť, chudoba, obmedzený prístup k zdravotnej starostlivosti a diskriminácia, je úloha sociálnych pracovníkov nevyhnutná a nenahraditeľná.

Jedným z najdôležitejších aspektov ich práce je pôsobenie ako sprostredkovateľov medzi komunitou a verejnými inštitúciami. Pomáhajú prekonávať jazykové, kultúrne a administratívne bariéry a slúžia ako most medzi obyvateľmi MRK a systémom sociálnych služieb, vzdelávania, zdravotnej starostlivosti alebo miestnej samosprávy. Rovnako dôležité je aj dlhodobé budovanie dôvery v komunitách, ktoré je predpokladom pozitívnej zmeny postojov a správania. Dôvera v sociálneho pracovníka môže predstavovať prvý krok k prekonaniu rezignácie a otvoreniu sa novým príležitostiam vedúcim ku kvalitnejšiemu životu jednotlivcov aj komunity ako celku.

V kontexte zberu údajov sociálni pracovníci využívajú túto vybudovanú dôveru a zabezpečujú kvalitu vstupných údajov (napr. znalosť poberania sociálnych dávok

a ich výšky). Sú odborne školení Štatistickým úradom Slovenskej republiky vo všetkých metodologických aspektoch zisťovania. Každému sociálnemu pracovníkovi je pridelený supervízor zodpovedný za metodické dodržiavanie postupov a poskytovanie podpory v prípade potreby.

Skúsenosti naznačujú, že táto kombinácia – sociálny pracovník ako anketár a metodický expert zo ŠÚ SR ako supervízor – vedie k vysokej miere odpovedí a kvalite údajov, ktorá minimalizuje potrebu imputácií. Na rozdiel od zisťovania EU SILC, kde miera imputácie príjmovej premennej dosahuje približne 20 %, je miera položkovej neodpovede minimálna. Podobný trend možno pozorovať aj pri celkovej jednotkovej miere odpovedí jednotlivých domácností, kde rozdiel medzi SILC MRK 2024 a novým panelom EU SILC 2024 dosahuje úroveň 14 % (72 % oproti 58 %).

3. REFERENČNÁ POPULÁCIA

Referenčnú populáciu zisťovania SILC MRK tvoria všetky osoby s obvyklým pobytom v domácnostiach nachádzajúcich sa v oblastiach marginalizovaných rómskych komunít. Výber je ďalej rozšírený aj na sociálne integrovaných Rómov.

Marginalizované rómske komunity sa považujú za sídelné oblasti s vysokou koncentráciou Rómov. Ich obyvatelia majú obmedzený prístup k verejným službám, tovarom, zdrojom a participácii na miestnom rozhodovaní. Zároveň majú obmedzený kontakt s ostatnými obyvateľmi obce žijúcimi mimo osady. Tieto osady sa nachádzajú na nižšej socioekonomickej úrovni v porovnaní s ostatnými časťami obce a vykazujú špecifické charakteristiky, ako sú nižšia kvalita bývania a preľudnenosť, nedostatok infraštruktúry, vysoká nezamestnanosť, príjmová aj materiálna chudoba, nižšia úroveň vzdelania a odlišná veková štruktúra (s vyšším podielom mladších vekových skupín).

Na účely tohto článku budeme jadrom obce označovať hlavnú, súvislú a zvyčajne najhustejšie obývanú časť obce, kde žije väčšina obyvateľov. Typicky zahŕňa administratívne centrum a hlavné obytné zóny, pričom vylučuje periférne oblasti, okraje, vzdialené osady a izolované lokality.

Na účely zisťovania SILC MRK sa marginalizované rómske komunity rozdeľujú do štyroch navzájom disjunktných tried, klasifikovaných podľa typu osídlenia. Tieto triedy sú definované v Atlase rómskych komunít:

1. koncentrácie na okraji obce – domácnosti žijúce koncentrovane oddelene na okraji obce, pričom existuje istá fyzická alebo sociálna hranica oddeľujúca tieto časti od väčšinovej populácie,
2. koncentrácie vo vnútri obce (vrátane jadier obcí s viac ako 30 % a najviac 50 % rómskeho obyvateľstva) – domácnosti žijúce koncentrovane integrovane vo vnútri obce, pričom existuje hranica oddeľujúca tieto domácnosti od väčšinovej populácie,
3. segregované osídlenia – domácnosti žijúce v osadách, ktoré sú vzdialené od obce alebo od nej oddelené prirodzenou či umelou bariérou (napr. les, cesta, atď.),
4. obce s rómskou väčšinou (viac ako 50 % obyvateľov) v jadre obce.

Na účely zisťovania SILC MKR musí osídlenie zahŕňať najmenej 30 obyvateľov alebo aspoň 5 domácností, aby mohlo byť klasifikované ako marginalizovaná rómska komunita. Ako sme uviedli, zisťovanie pokrýva aj skupinu sociálne integrovaných

Rómov, ktorí nepatria do žiadnej z uvedených tried. Ide o domácnosti, ktoré žijú v rámci obce zmiešane s majoritnou populáciou.

4. OPORTA VÝBERU

Asi najznámejší zdroj dát o Rómoch je Atlas rómskych komunít. Úrad splnomocnenca vlády SR pre rómske komunity (2019) uvádza, že atlas je nástrojom na mapovanie štrukturálnych nerovností medzi Rómami a majoritou. Vo výsledku obsahuje širokú paletu informácií o obciach s rómskymi osídleniami a o samotných osídleniach, preto je dobrým východiskovým zdrojom pre nápravu nerovností prostredníctvom verejných politík. Jeho hlavným cieľom je komplexne zmapovať rómske komunity na území Slovenskej republiky – predovšetkým tie, ktoré žijú v podmienkach sociálneho vylúčenia a priestorovej marginalizácie. Ako uvádza Škobla (2025), projekt vytvára unikátnu databázu a priestorovú mapu rómskych osídlení, ktorá slúži ako empirický základ na tvorbu, plánovanie a hodnotenie verejných politík v oblasti sociálnej inklúzie Rómov. Vo výsledku obsahuje širokú paletu informácií o samotných osídleniach, preto je dobrým východiskovým zdrojom pri náprave nerovností prostredníctvom verejných politík.

Atlas zhromažďuje a spracúva údaje, ktoré o osídleniach a ich obyvateľoch poskytujú jednotlivé samosprávy a ich predstavitelia. Nezberá tak sebadeklaratórne individuálne osobné údaje, ale výhradne štrukturálne, agregované charakteristiky rómskej populácie a jej priestorového rozmiestnenia. Príslušnosť k rómskej komunite je de facto pripísaná všetkým obyvateľom uvedenej koncentrácie. Ravasz et al. (2020) upozorňuje, že pri uvedených počtoch obyvateľov jednotlivých koncentrácií ide o celkové odhadované počty ľudí, ktorí dlhodobo žijú v danej koncentrácii, nie o počet ľudí, ktorí sa hlásia k rómskej menšine. Ďalej uvádza, že takto územne definované východisko ku skúmaniu životných podmienok Rómov je založené na predpoklade, že veľká časť rómskej populácie žije v určitých priestorových celkoch – či už vo vnútri obcí, na ich okraji alebo v segregovaných osídleniach, a na základe toho je možné ich mapovanie. Podľa Škoblu (2025), atlas možno teda vnímať predovšetkým ako mapovanie osídlení rómskych komunít a ich priestorového rozloženia v rámci Slovenska. Vychádza z predpokladu, že koncentrovane žijúce rómske komunity sa vo väčšine prípadov viažu na konkrétne lokality vo vnútri, na okraji alebo mimo obce. Tieto lokality zároveň spravidla vykazujú znaky štrukturálnej marginalizácie.

Detailné informácie atlasu o koncentrovane žijúcich rómskych komunitách a ich geografickom umiestnení by mohli poskytnúť vhodný východiskový bod nielen na definíciu referenčnej populácie zisťovania SILC MRK, ale aj na zostrojenie opory výberu. Ako poznamenáva Škobla (2025), tento prístup, hoci prakticky účelný z hľadiska terénnej organizácie výskumu, prináša riziko tzv. etnickej atribúcie zvonku a zároveň vedie k podhodnoteniu diverzity rómskej populácie. Absentuje tým aj možnosť preskúmať subjektívnu dimenziu etnicity – teda ako sa samotní respondenti identifikujú a ako táto identita ovplyvňuje ich sociálne postavenie a skúsenosť so začlenením.

Navyše na vytvorenie reprezentatívnej výberovej vzorky zisťovania SILC MRK formou náhodného pravdepodobnostného výberu je výhradné použitie Atlasu rómskych komunít nepostačujúce. Pomocou neho je síce možné kategorizovať jednotlivé obce a ich časti do zadefinovaných skupín, no bez príslušnej databázy na úrovni jednotlivých domácností je zber údajov obmedzený na realizáciu náhodnej

prechádzky alebo kvótného výberu. Oba spôsoby však z hľadiska zabezpečenia vysokej kvality štatistického zisťovania predstavujú určité riziká – v prvom prípade vo forme ohrozenia úplnej náhodnosti výberu, v druhom prípade vnášaním možného vychýlenia do jednotlivých odhadov.

Databáza obývaných bytových domácností zo Sčítania obyvateľov, domov a bytov 2021 predstavuje spoľahlivý zdroj na realizáciu pravdepodobnostných náhodných výberov všetkých zisťovaní v domácnostiach realizovaných Štatistickým úradom Slovenskej republiky. Na účely zisťovania SILC MRK v nej však absentujú kľúčové informácie o type osídlenia marginalizovaných rómskych komunít. Z toho dôvodu je vhodné vykonať geopriestorovú zhľukovú analýzu, ktorá má za cieľ priradiť jednotlivým domácnostiam MRK pravdepodobnú kategorizáciu. Takto upravená databáza, navyše v spojení s Atlasom rómskych komunít 2019, predstavuje možnú oporu výberu, ktorá umožní vytvorenie reprezentatívnej vzorky zisťovania SILC MRK na báze náhodného výberu.

V prvom kroku sa z SODB 2021 pripraví databáza bytových domácností obývaných aspoň jedným Rómom. Pri pohľade na výsledky sčítania môžeme pozorovať diskrepanciu medzi počtom obyvateľov rómskej národnosti a odhadmi publikovanými v Atlase rómskych komunít. Dôvodom môže byť rozdielny koncept oboch ukazovateľov – zatiaľ čo v prvom prípade ide o subjektívne vnímanú etnicitu, v druhom prípade sa zohľadňuje vonkajšia identifikácia marginalizovanej komunity. Vzhľadom na to, že aj osoby s inou subjektívne deklarovanou etnicitou než rómskou, ktoré sú súčasťou marginalizovaných rómskych komunít, čelia v porovnaní s väčšinou populáciou obmedzeným životným podmienkam komunity, a zároveň vzhľadom na potrebu metodického priblíženia zisťovania s atlasom, budeme výhradne na účely zostrojenia opory výberu považovať domácnosti s prítomnosťou Rómov za rómske domácnosti. S cieľom maximalizácie detekcie potenciálnych marginalizovaných komunít v rámci priestorovej analýzy sa osoba bude považovať za Róma, ak spĺňa aspoň jedno z nasledujúcich kritérií:

- prvá národnosť rómska,
- druhá národnosť rómska,
- materinský jazyk rómsky.

Celkový počet takto definovaných osôb so súčasným pobytom v SR predstavuje približne 170 000, pričom celkový počet všetkých osôb žijúcich vo vyššie definovaných rómskych domácnostiach dosahuje približne 274 000. Tieto osoby obývajú spolu viac ako 36 600 bytov, z ktorých

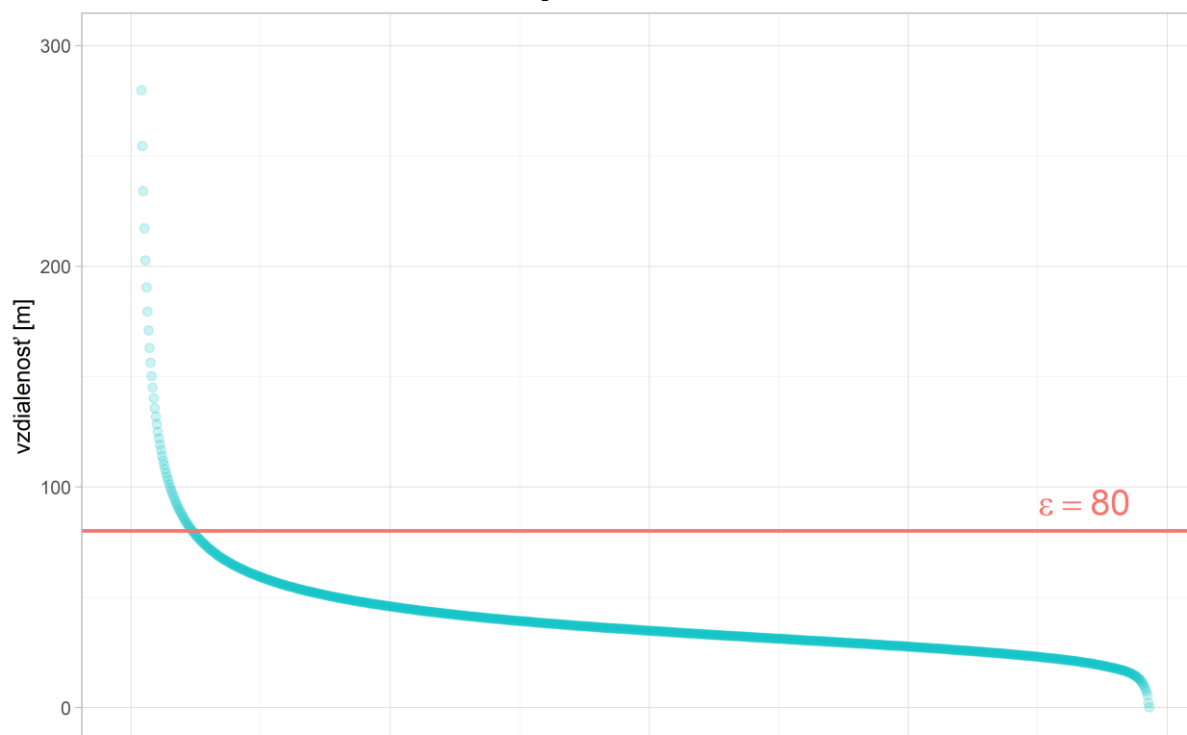
- 97,3 % môže byť považovaných za súkromné domácnosti,
- 1,2 % spadá medzi ostatné kategórie bytov, najmä medzi kolektívne domácnosti,
- 1,5 % sú centroidy obcí a obecné úrady.

V druhom kroku sa všetky domácnosti obývané podľa súčasného pobytu rozdelia v obci do jednotlivých zhľukov (klastrov) podľa vzájomnej pozície a vzdialenosti. Izolované alebo samostatne ležiace adresné body predstavujú odľahlé pozorovania (šum), ktoré nie sú priradené k žiadnemu klastru. Na tento účel sa využije algoritmus DBSCAN (Ester et al., 1996), ktorý je jedným z najrozšírenejších nástrojov na zhľukovú analýzu v geopriestorových databázach. Jeho princípom je klasifikácia bodov

v priestore do jednotlivých tried. Využíva dva nevyhnutné parametre – kruhovú vzdialenosť okolo sledovaného bodu (ε -okolie) a minimálny počet bodov v ε -okolí na označenie daného bodu ako jadrového (*minPts*). Klastrom je zoskupenie všetkých navzájom hustotne dosiahnuteľných jadrových bodov a všetkých k nim prislúchajúcich hustotne prepojených hraničných bodov. Body, ktoré nie sú v ε -okolí žiadneho jadrového bodu, sú odľahlými a nie sú súčasťou žiadneho z klastrov. Na zefektívnenie chodu algoritmu uprednostníme pri klasifikácii domácností do zhlukov úroveň adresných bodov pred úrovňou bytových domácností.

Voľba vhodných parametrov algoritmu závisí od priestorového rozptýlenia bodov databázy. Vzhľadom na to, že hustota osídlenia jednotlivých obcí je veľmi rôznorodá, nastavenie jednotných parametrov nie je z pohľadu očakávaných výstupov prijateľné. Ester et al. (1996) vo svojej práci neodporúčajú pre 2-rozmerné databázy voľbu hodnoty *minPts* vyššiu ako 4, pretože by to viedlo k výraznému zvýšeniu výpočtových nárokov bez signifikantného zlepšenia výsledkov. Na základe testovania algoritmu na údajoch SODB 2021 sme na tvorbu opory výberu zisťovania SILC MRK zvolili *minPts* = 3 pre všetky obce. Na určenie optimálneho polomeru ε -okolía Ester et al. (1996) odporúčajú využiť graf zoradených vzdialeností ku k -temu najbližšiemu susedovi (v našom prípade $k = 3$), ktorý zobrazuje vzdialenosti všetkých bodov databázy. Hodnota polomeru ε sa následne určí vizuálnou identifikáciou centrálného bodu pozorovateľného „laktá“ krivky. Táto hranica oddeľuje jadrové a hraničné body klastrov od odľahlých pozorovaní tvoriacich šum. Pomocou grafu č. 1 nastavíme aplikáciou odporúčaného postupu hodnotu polomeru okolia jadrových bodov na $\varepsilon = 80$ metrov.

Graf č. 1: Zoradené vzdialenosti k 3. najbližšiemu susedovi

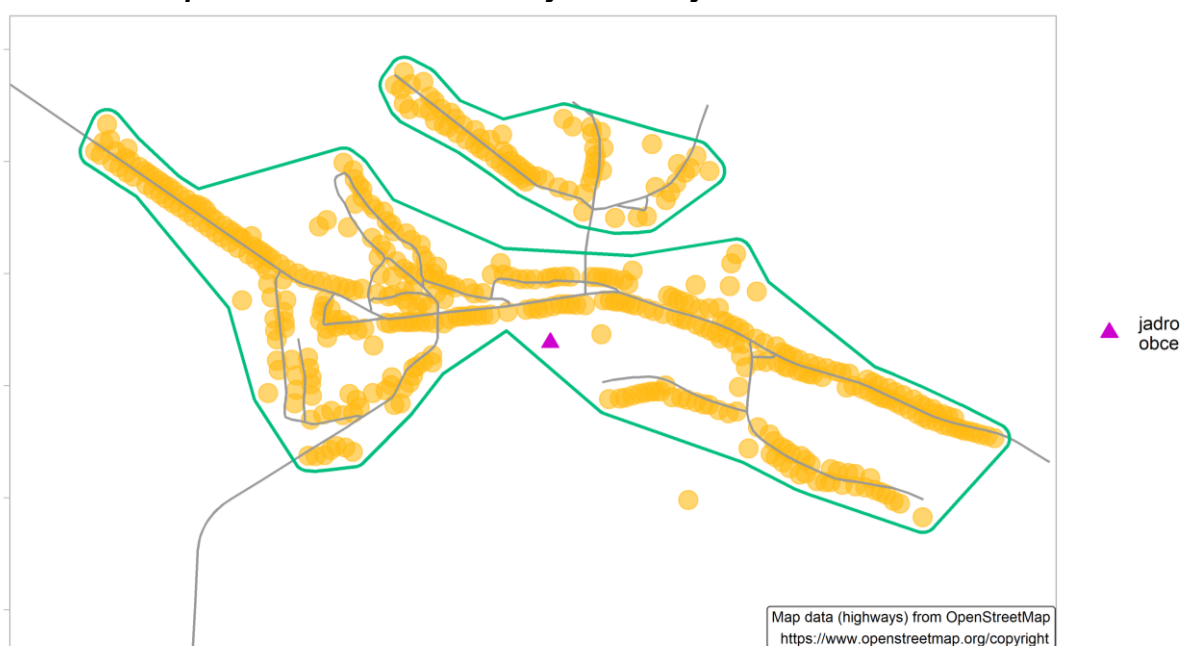


Zdroj údajov: Štatistický úrad Slovenskej republiky (2021); výpočty autorov

Zvolený parameter sa ukazuje ako prijateľný v prípade obcí s priemernou a zároveň stabilnou hustotou obývanosti v rámci celého ich územia, čo sa demonštruje na grafe č. 2. Môžeme pozorovať vytvorenie dvoch klastrov, pričom väčší z nich možno na základe vizuálneho posúdenia identifikovať ako pravdepodobné jadro obce a menší ako pridruženú sekundárnu zastavanú lokalitu. Na základe analýzy vykonanej na vzorke všetkých obcí SR je klaster automatizovane označený ako jadro obce, ak je splnená niektorá z nasledujúcich podmienok (zorađených postupne podľa priorít):

- v klasteri sa nachádza obecný úrad a ide zároveň o najväčší klaster z hľadiska rozlohy alebo počtu obyvateľov,
- klaster je najväčší z hľadiska hodnoty $\frac{pop_i}{\max(pop)} * \frac{area_i}{\max(area)}$, kde pop_i označuje počet obyvateľov klastra i a $area_i$ jeho rozlohu.

Graf č. 2: Geopriestorová zhluková analýza menšej obce – identifikácia klastrov



Zdroj údajov: Štatistický úrad Slovenskej republiky (2021); výpočty autorov

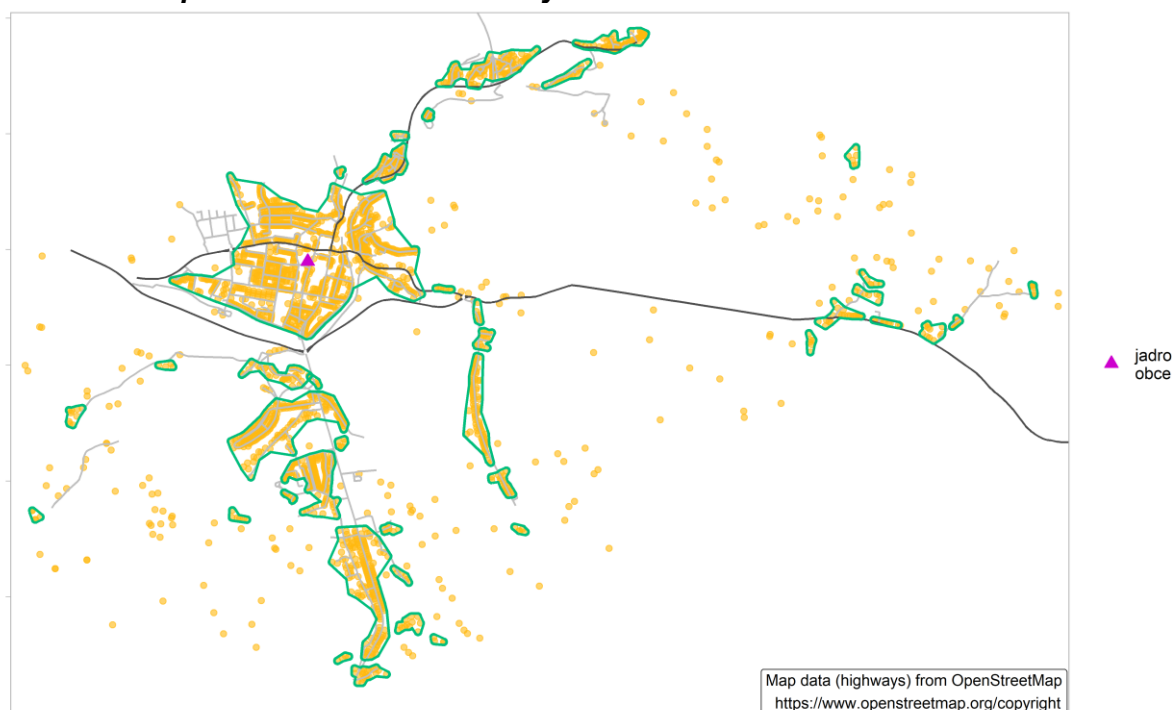
Vzhľadom na výraznú diverzitu v hustote obývanosti medzi jednotlivými obcami nie je zvolená hodnota vyhovujúca ani pre riedko osídlené obce, ani pre obce s výrazne heterogénnou hustotou osídlenia v rámci ich územia (najmä prípad miest). Graf č. 3 demonštruje realizáciu zhlukovej analýzy vo väčšom meste s fixne nastaveným parametrom $\varepsilon = 80$. Pozorujeme vznik vysokého počtu zhlukov reprezentujúcich menšie sídelné útvary spolu s vysokým objemom šumu, čo nie je z hľadiska cieľov analýzy žiaduce. Pri konštrukcii opory výberu s osobitným zameraním na marginalizované rómske komunity je kľúčové jednoznačne definovať jadro obce pri súčasnom obmedzení počtu sekundárnych klastrov, ktoré by mohli napomôcť lokalizácii segregovaných osád. Počet odľahlých nezatriedených bodov by mal byť minimálny a obmedzený iba na skutočný šum, keďže práve na okrajových adresných bodoch by mohli byť lokalizované marginalizované koncentrácie. Nastavenie maximálneho počtu klastrov v obci sa však neukázalo ako optimálne. Na základe výsledkov analýz preto aplikujeme dva adaptívne mechanizmy na konštrukciu klastrov na úrovni jednotlivých obcí s cieľom vytvoriť univerzálne riešenie:

1. Dynamicky detegovaná hodnota parametra ε – počiatočne sa nastaví $\varepsilon = 80$ a v jednotlivých krokoch sa postupne zvyšuje, až kým nie je celkový objem šumu nižší ako 2 %. Táto hranica sa v analýze na vzorke obcí ukázala ako vhodná na minimalizáciu objemu takého šumu, ktorý v skutočnosti predstavuje geograficky blízke domácnosti tvoriace ohraničené osídlenia, a súčasne na zamedzenie vytvárania klastrov z priestorovo rozptýlených domácností na okrajoch obcí (predstavujúcich skutočný šum). Výsledný polomer tak reflektuje skutočnú hustotu osídlenia obce so súčasnou minimalizáciou klasifikácie hraničných bodov ako odľahlých.
2. Zlučovanie navzájom blízky klastrov, t. j. klastrov, ktorých najmenšia vzájomná vzdialenosť je nižšia ako

$$\varepsilon \left(\sqrt[11]{N_m} - 1 \right),$$

kde N_m predstavuje počet obyvateľov obce m podľa súčasného pobytu.

Graf č. 3: Geopriestorová zhluková analýza väčšieho mesta – identifikácia klastrov



Zdroj údajov: Štatistický úrad Slovenskej republiky (2021); výpočty autorov

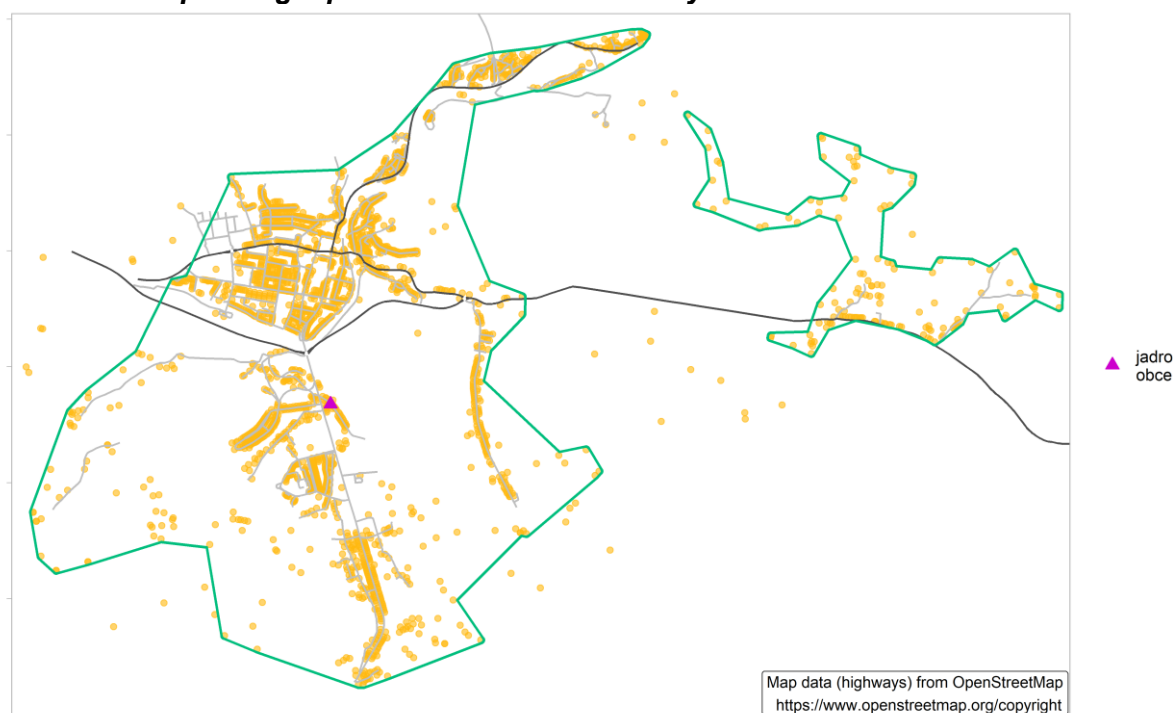
Graf č. 4 zobrazuje realizáciu geopriestorovej zhlukovej analýzy na území daného mesta využitím oboch adaptívnych mechanizmov. Môžeme pozorovať ich jasný vplyv vzhľadom na nastavené ciele. Na prvom grafe vysoký počet zhlukov znemožňuje rozlíšenie medzi vnútornými a okrajovými časťami, zatiaľ čo na druhom obrázku môžeme vidieť zachytenú priestorovú štruktúru mesta z hľadiska obývaných oblastí.

Uskutočnenie geopriestorovej zhlukovej analýzy všetkých obcí je predpokladom na detekciu referenčných domácností zisťovania SILC MRK a ich klasifikáciu do jednotlivých kategórií. Na tento účel sa rovnako využije zhluková analýza. V tomto prípade sa držíme predvolenej hodnoty $\varepsilon = 80$ metrov, keďže sa zameriavame na oblasti s vysokou koncentráciou Rómov, a preto chceme zabrániť vzniku riedkych

klastrov. Na zníženie možného vysokého počtu MRK v obciach sme použili algoritmus zlučovania blízkych klastrov. Podmienkou na označenie klastra ako MRK je úhrnná prítomnosť minimálne 30 osôb v aspoň štyroch rómskych domácnostiach. Domácnosti, ktoré nie sú klasifikované do žiadneho klastra MRK, sa na účely opory výberu považujú za priestorovo integrované.

Graf č. 5 znázorňuje klasifikáciu viacerých marginalizovaných rómskych komunití a priestorovo integrovaných domácností v náhodnej obci. Všetky rómske domácnosti boli generované ako syntetické dáta a nezodpovedajú žiadnym skutočným osobám ani domácnostiam.

Graf č. 4: Adaptívna geopriestorová zhluková analýza väčšieho mesta



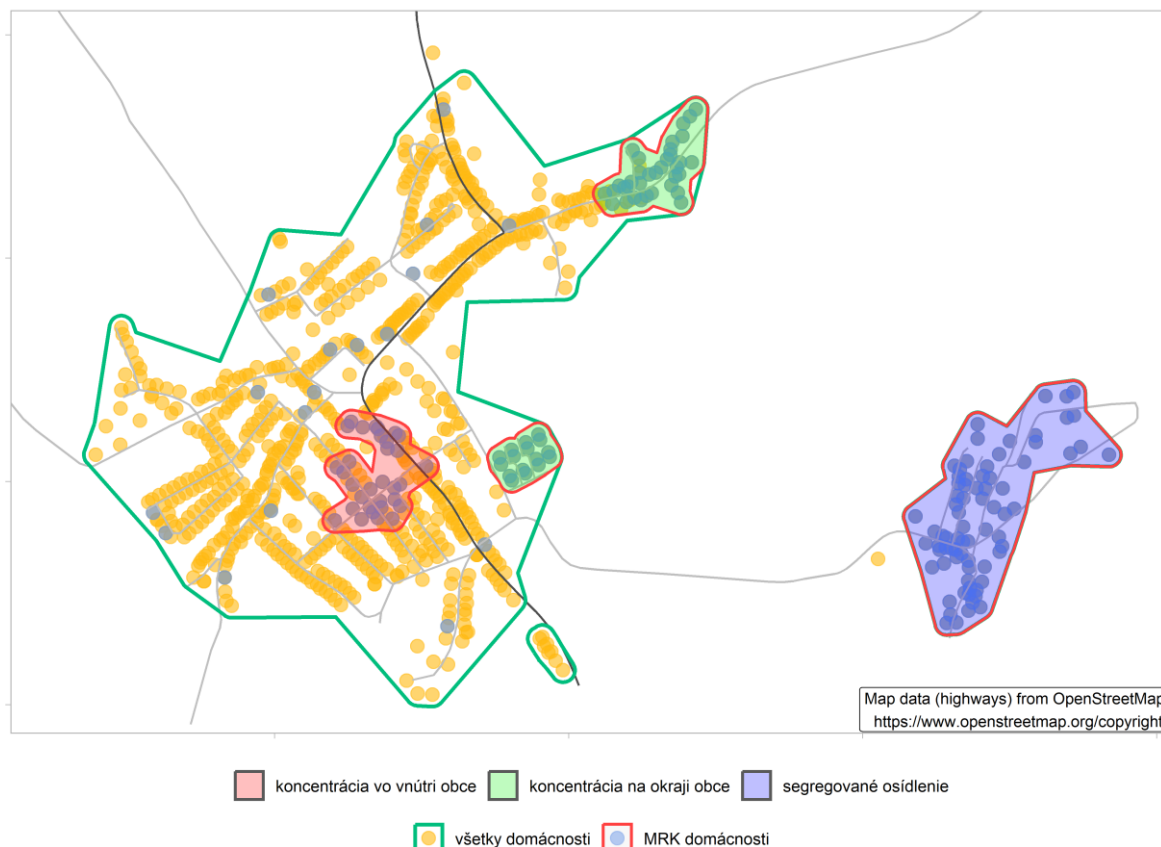
Zdroj údajov: Štatistický úrad Slovenskej republiky (2021); výpočty autorov

Detegované komunity sú automaticky zatriedené do jednotlivých kategórií, ktorých definície sú do veľkej miery analogické s triedami podľa typu osídlenia vzťahujúcimi sa na referenčnú populáciu:

1. *rozptýlení v jadre* – integrované rómske domácnosti nachádzajúce sa v jadre obce mimo klastra MRK,
2. *koncentrácia vo vnútri jadra* – rómske domácnosti nachádzajúce sa v klasteri MRK ležiacom vo vnútri jadra obce,
3. *koncentrácia na okraji jadra* – rómske domácnosti nachádzajúce sa v klasteri MRK ležiacom na okraji jadra obce,
4. *rozptýlení mimo jadra* – rómske domácnosti nachádzajúce sa mimo jadra obce i mimo klastra MRK,
5. *koncentrácia mimo jadra* – rómske domácnosti nachádzajúce sa v klasteri MRK ležiacom mimo jadra obce, ktorý je podmnožinou väčšieho klastra vytvoreného zo všetkých domov; podiel počtu osôb rómskych domácností a celkového počtu osôb vo väčšom klasteri sa rovná maximálne 80 %,

6. *segregovaná koncentrácia* – rómske domácnosti nachádzajúce sa v klastrí ležiacom mimo jadra obce; podiel počtu osôb rómskych domácností a celkového počtu osôb v klastrí sa rovná minimálne 80 %,
7. *obec s rómskou majoritou* – rómske domácnosti v obci, v ktorej žije v rómskych domácnostiach ležiacich v jadre (mimo klastrov na okraji) viac ako 50 % celkovej populácie jadra obce.

Graf č. 5: Klasifikácia domácností MRK v náhodnej obci



Zdroj údajov: Štatistický úrad Slovenskej republiky (2021), syntetické údaje; výpočty autorov

Klasifikácia lokalít do jednotlivých kategórií je iba indikatívna a nie je totožná s ich zatriedením podľa Atlasu rómskych komunít 2019. Dôvody rozdielov spočívajú:

- v rozdielnej definícii rómskej populácie medzi SODB a atlasom,
- v subjektívnom posudzovaní lokalít predstaviteľmi samospráv – ide o presnú metódu zatriedenia, ktorá zohľadňuje miestne reálie, marginalizáciu komunity i socioekonomické charakteristiky, ktoré nie je možné zachytiť algoritmi pri priestorovej zhlukovej analýze,
- v rozdielnom počte detegovaných a skutočných klastrov v jednotlivých obciach.

Jeden z rozdielov je možné pozorovať v počte obyvateľov koncentrácií vo vnútri jadier obcí. Viaceré z nich sú totiž vzhľadom na svoju polohu vo vzťahu k jadrovému klastru algoritmicky klasifikované ako koncentrácie na okraji, hoci sú na základe miestnych skutočností majoritou vnímané ako vnútroobecné. Problematická je aj množina koncentrácií mimo jadra, ktoré z dôvodu nižšej hustoty rómskych domácností

nie sú označené ako segregované. Lokality tejto kategórie však v skutočnosti vo veľkej miere patria práve do skupiny segregovaných osád, ako aj medzi koncentrácie na okraji alebo vo vnútri obcí.

Na porovnanie klasifikácie lokalít vykonanej zhlukovou analýzou na údajoch SODB so zatriedením podľa atlasu sme použili databázu Atlasu rómskych komunít 2019, ktorú zverejňuje Inštitút Mateja Bela². Hoci nejde o oficiálnu databázu atlasu (v prípade niektorých obcí s viac ako 30 % rómskej populácie v jadre sú príslušné osídlenia vnútri obcí charakterizované ako integrované), predstavuje užitočný zdroj informácií o výsledkoch mapovania sídiel rómskych komunít. S cieľom zosúladiť jednotlivé ukazovatele s definíciou referenčnej populácie podľa metodiky zisťovania SILC MRK (ktorá pracuje aj s kategóriou *obec s rómskou majoritou*) sme prepočítali celkové objemy populácie a počty osídlení podľa typu.

Tabuľka č. 1 zobrazuje počty automaticky detegovaných rómskych koncentrácií porovnané s prepočítanými údajmi atlasu. Pri údajoch sčítania obyvateľov možno z dôvodov uvedených predtým pozorovať nižší celkový objem detegovaných koncentrácií, ako aj mierne vychýlenie v zastúpení koncentrácií nachádzajúcich sa vo vnútri a mimo obcí.

Tabuľka č. 1: Porovnanie odhadovaných počtov rómskych koncentrácií

	SODB 2021		ARK 2019	
koncentrácia vo vnútri obce	216	27 %	350	33 %
koncentrácia na okraji obce	359	46 %	502	48 %
koncentrácia mimo obce (segregovaná)	213	27 %	194	19 %
obec s rómskou väčšinou v jadre	36		85	

Zdroj: Štatistický úrad Slovenskej republiky (2021), Inštitút Mateja Bela (2019); výpočty autorov

5. VÝBEROVÁ VZORKA

Konštrukcia opory výberu umožňuje vytvorenie reprezentatívnej výberovej vzorky metódou pravdepodobnostného náhodného výberu domácností. Stratifikácia je na úrovni samosprávnych krajov a typov osídlení. Na zefektívnenie zberu údajov v teréne a dosiahnutie maximálnej možnej zhody medzi definíciou referenčnej populácie zisťovania a pokrytím samotnej vzorky sme do nej vybrali iba domácnosti z obcí, ktoré sa v klasifikácii typu osídlenia zhodovali medzi Atlasom rómskych komunít 2019 a výsledkami geopriestorovej zhlukovej analýzy na údajoch SODB 2021. Čistú veľkosť výberu sme stanovili na 1 000 domácností pre marginalizované rómske komunity a 250 domácností pre priestorovo integrovanú populáciu. Alokácia vzorky do skupín MRK podľa typu osídlenia je približne proporcionálne úmerná odhadom ich veľkosti na základe dvoch posledných atlasov.

Pri porovnaní celkových odhadovaných počtov obyvateľov jednotlivých typov osídlení podľa atlasu a podľa zhlukovej analýzy (v druhom prípade sme do celkových odhadov započítali všetky osoby žijúce v koncentracii, nielen Rómovia) spôsobuje najväčší rozdiel problematická množina *koncentrácií mimo jadra*. Ide o kategóriu, ktorá

² <https://www.institutmatejabela.sk/ark2019>

bola v geopriestorovej zhlukovej analýze vyhodnotená ako lokalita ležiaca mimo jadra obce nevykazujúca zrejme znaky segregovaného osídlenia (najmä z hľadiska hustoty osídlenia). Je pravdepodobné, že časť domácností tejto kategórie prislúcha ku skupine segregovaných osídlení a časť ku koncentráciám na okraji obce. Tabuľky č. 2 a 3 porovnávajú odhadované objemy populácií v jednotlivých typoch triedenia. V prvom prípade sú koncentrácie mimo jadra klasifikované ako segregované osídlenia, zatiaľ čo v druhom prípade ako koncentrácie na okraji obcí. Napriek rozdielom v celkových úhrnoch (rozoberaným v 4. kapitole) je relevantné sledovať percentuálne zastúpenie jednotlivých skupín referenčnej populácie.

Tabuľka č. 2: Porovnanie celkových odhadovaných počtov obyvateľov jednotlivých typov osídlení – koncentrácia mimo jadra zlúčená s koncentráciou mimo obce

	SODB 2021		ARK 2019	
koncentrácia vo vnútri obce	22 340	8 %	50 092	12 %
koncentrácia na okraji obce	73 976	27 %	151 410	36 %
koncentrácia mimo obce (segregovaná)	71 723	26 %	56 365	14 %
obec s rómskou väčšinou v jadre	19 004	7 %	50 388	12 %
priestorová integrácia	86 655	32 %	109 280	26 %

Zdroj: Štatistický úrad Slovenskej republiky (2021), Inštitút Mateja Bela (2019); výpočty autorov

Tabuľka č. 3: Porovnanie celkových odhadovaných počtov obyvateľov jednotlivých typov osídlení – koncentrácia mimo jadra zlúčená s koncentráciou na okraji obce

	SODB 2021		ARK 2019	
koncentrácia vo vnútri obce	22 340	8 %	50 092	12 %
koncentrácia na okraji obce	100 431	37 %	151 410	36 %
koncentrácia mimo obce (segregovaná)	45 268	16 %	56 365	14 %
obec s rómskou väčšinou v jadre	19 004	7 %	50 388	12 %
priestorová integrácia	86 655	32 %	109 280	26 %

Zdroj: Štatistický úrad Slovenskej republiky (2021), Inštitút Mateja Bela (2019); výpočty autorov

Tabuľka č. 4: Čistá veľkosť výberovej vzorky – počet domácností

	VEĽKOSŤ VZORKY	
koncentrácia vo vnútri obce	170	14 %
koncentrácia na okraji obce	555	44 %
koncentrácia mimo obce (segregovaná)	180	14 %
obec s rómskou väčšinou v jadre	95	8 %
priestorová integrácia	250	20 %

Zdroj: výpočty autorov

6. ZÁVER

Predpokladom zabezpečenia vysokej kvality štatistického zisťovania SILC MRK je popri vyriešení veľkého množstva metodických a technických otázok aj vytvorenie reprezentatívnej výberovej vzorky metódou pravdepodobnostného náhodného výberu, ktorá umožňuje zovšeobecnenie výsledkov na celú populáciu. Na jeho realizáciu je nevyhnutné disponovať oporou výberu obsahujúcou kontaktné údaje jednotlivých spravodajských jednotiek spolu s nevyhnutnými stratifikačnými premennými, pričom táto opora musí čo najpresnejšie pokrývať celú referenčnú populáciu. V prípade zisťovania SILC MRK ju tvoria všetky domácnosti marginalizovaných rómskych komunít podľa jednotlivých typov osídlenia.

Najpresnejším zdrojom údajov o danej populácii je Atlas rómskych komunít, ktorý však nezbiera individuálne údaje, ale výhradne agregované charakteristiky rómskej populácie. Štandardnou oporou výberu pri všetkých zisťovaniach v domácnostiach, ktoré vykonáva ŠÚ SR, je databáza obývaných bytov vytvorená na základe SODB 2021. Na účely zisťovania SILC MRK v nej však absentujú kľúčové informácie o priestorovom rozmiestnení rómskej populácie v jednotlivých lokalitách. Tieto údaje boli do databázy imputované prostredníctvom geopriestorovej zhľukovej analýzy.

Porovnanie výsledkov analýzy s odhadmi Atlasu rómskych komunít 2019 indikuje približné zosúladenie na úrovni podielov jednotlivých kategórií v rámci celej množiny MRK. Prepojenie opory výberu so štrukturálnymi charakteristikami atlasu následne umožnilo vytvorenie reprezentatívnej výberovej vzorky.

LITERATÚRA

- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, & U. Fayyad (Eds.), *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)* (pp. 226 – 231). AAAI Press.
- Inštitút Mateja Bela. (2019). *Databáza ARK 2019* [Data set].
https://www.institutmatejabela.sk/files/ugd/1a16af_6d8447b32414462e8e528dcc627d2e34.xlsx
- Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal*, 10(1), 439 – 446. <https://doi.org/10.32614/RJ-2018-009>
- Pebesma, E., & Bivand, R. (2023). *Spatial data science: With applications in R*. Chapman and Hall/CRC. <https://doi.org/10.1201/9780429459016>
- R Core Team. (2026). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ravasz, Á., et al. (2020). *Atlas rómskych komunít 2019*. Inštitút Mateja Bela.
https://www.institutmatejabela.sk/files/ugd/1a16af_7b05b19393f749f9bd37451422dc68e3.pdf
- Škobla, D. (2025). Zdroje údajov o životných podmienkach Rómov a ako týmto údajom rozumieť? Metodologická poznámka. *Bulletin IVPR*. Inštitút pre výskum práce a rodiny.
- Štatistický úrad Slovenskej republiky. (2021). *Sčítanie obyvateľov, domov a bytov 2021*. <https://www.scitanie.sk/>
- Úrad splnomocnenca vlády SR pre rómske komunity. (2019). *Atlas rómskych komunít 2019*.
<https://romovia.vlada.gov.sk/atlas-romskych-komunit/atlas-romskych-komunit-2019/>

Úrad splnomocnenca vlády SR pre rómske komunity. (2021). *Stratégia rovnosti, inklúzie a participácie Rómov do roku 2030*.

https://romovia.vlada.gov.sk/site/assets/files/11113/strategia_2030.pdf

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), Article 1686. <https://doi.org/10.21105/joss.01686>

RESUMÉ

Článok sa zaoberá problematikou vytvorenia reprezentatívnej výberovej vzorky zisťovania SILC MRK, ktoré slúži na kvantifikáciu chudoby a sociálneho vylúčenia v prostredí marginalizovaných rómskych komunít. Hlavným cieľom bolo vytvoriť oporu výberu umožňujúcu realizáciu pravdepodobnostného náhodného výberu domácností pri súčasnom čo najpresnejšom pokrytí referenčnej populácie podľa typov osídlení definovaných v Atlase rómskych komunít.

V článku je predstavený postup využitia metódy geopriestorovej zhlukovej analýzy pomocou algoritmu DBSCAN na automatizovanú detekciu lokalít rómskych osídlení v databáze obývaných bytov vytvorenej z údajov Sčítania obyvateľov, domov a bytov 2021. Prostredníctvom adaptívnych mechanizmov klasifikácie boli identifikované a kategorizované jednotlivé koncentrácie vrátane segregovaných osídlení a priestorovo integrovaných domácností. Výsledky analýzy sa následne porovnali s údajmi Atlasu rómskych komunít 2019.

Navrhnutý metodický prístup umožnil rozšírenie databázy domácností o údaje nevyhnutné na vytvorenie reprezentatívnej výberovej vzorky štatistického zisťovania SILC MRK, ktorá vytvára predpoklady na produkciu kvalitných a medzinárodne porovnateľných ukazovateľov o životných podmienkach marginalizovaných rómskych komunít. Tie umožňujú efektívne monitorovanie sociálneho vylúčenia, chudoby a dopadov verejných politík zameraných na podporu inklúzie rómskej populácie.

Príspevok zároveň poukazuje na limity automatizovanej klasifikácie priestorových koncentrácií obyvateľstva, najmä vo vzťahu k subjektívne posudzovanej kategorizácii jednotlivých lokalít predstaviteľmi samospráv, ktorá je najpresnejším dostupným zdrojom relevantných údajov.

RESUME

The article addresses the issue of constructing a representative sample for the SILC MRK survey, which is aimed at measuring poverty and social exclusion within marginalised Roma communities. The main objective was to create a sampling frame enabling probability random sampling of households while ensuring the most accurate possible coverage of the reference population according to the settlement types defined in the Atlas of Roma Communities.

The paper presents a methodology for applying geospatial cluster analysis using the DBSCAN algorithm to automatically detect Roma settlement localities within the database of occupied dwellings derived from the 2021 Population and Housing Census. Through adaptive classification mechanisms, individual concentrations, including segregated settlements and spatially integrated households, were identified and categorised. The results of the analysis were subsequently compared with data from the 2019 Atlas of Roma Communities.

The proposed methodological approach enabled the extension of the household database with information necessary for constructing a representative sample for the SILC MRK survey, thereby creating the conditions for the production of high-quality and internationally comparable indicators on the living conditions of marginalised Roma communities. These indicators enable effective monitoring of social exclusion, poverty, and the impacts of public policies aimed at promoting the inclusion of the Roma population.

The paper also highlights the limitations of the automated classification of spatial population concentrations, particularly in relation to the subjectively assessed categorisation of individual localities by municipal representatives, which represents the most accurate available source of relevant information.

PROFESIJNÝ ŽIVOTOPIS

Mgr. Boris Frankovič je absolventom Fakulty matematiky, fyziky a informatiky Univerzity Komenského v Bratislave. Od roku 2010 pracuje v Štatistickom úrade Slovenskej republiky, kde sa venuje najmä tematike analýzy údajov, spracovania štatistických zisťovaní, ochrany dôverných štatistických údajov a realizácie školení. Spolupracuje taktiež so zahraničnými štatistickými úradmi a je stálym členom Expertnej skupiny pre ochranu dôverných štatistických údajov v Eurostate.

Ing. Monika Rákošová absolvovala inžinierske štúdium na Ekonomickej univerzite v Bratislave v odbore obchodný manažment (2012). Po ukončení štúdia pôsobila niekoľko rokov na alma mater v Centre podnikateľských činností a univerzitných služieb. Od roku 2020 pracuje v sekcii sociálnych štatistík a demografie Štatistického úradu Slovenskej republiky, kde sa v roku 2023 stala riaditeľkou odboru štatistiky životnej úrovne obyvateľstva. V tejto oblasti sa špecializuje na problematiku sociálnej ochrany podľa metodiky ESSPROS a výberové zisťovania v domácnostiach.

KONTAKT

boris.frankovic@statistics.sk

monika.rakosova@statistics.sk