

## Komparácia metód identifikácie determinantov životnej úrovne

Nad'a BIRČIAKOVÁ – Jana STÁVKOVÁ\* – Oldřich TRENZ\*\*

---

### Comparison of Methods for Identifying the Determinants of the Standard of Living

#### Abstract

*The article deals with the identification of factors of the standard of living. Ambiguous definition of the term “standard of living” requires for its quantification sufficient theoretical knowledge from the methodology field. The principal component analysis (PCA) was used to reduce the extensive amount of factors of the standard of living as well as for the determining of the most important ones. Data mining was used to compare the use of the PCA with various classification algorithms and later on it was tested with developments of the feature selection. The methods were applied on the sample of 2 783 respondents from 5 EU countries which represents areas of cultural similarities. The result of this effort is reduction from 99 considered factors of standard of living to final 45. Data mining helped to exclude 30 attributes and thus the final amount is set to 69. In case of comparison of these methods and their results it seems more appropriate the PCA over the feature selection method.*

**Keywords:** *factors, standard of living, principal component analysis, data mining, feature selection*

**JEL Classification:** I31, C80, H31

---

---

\* Nad'a BIRČIAKOVÁ – Jana STÁVKOVÁ, Mendelova univerzita v Brně, Provozně ekonomická fakulta, Ústav marketingu a obchodu, Zemědělská 1, 613 00 Brno, Česká republika; e-mail: nada.birciakova@mendelu.cz; jana.stavkova@mendelu.cz

\*\* Oldřich TRENZ, Mendelova univerzita v Brně, Provozně ekonomická fakulta, Ústav informatiky, Zemědělská 1, 613 00 Brno, Česká republika; e-mail: oldrich.trenz@mendelu.cz

**Pod'akovanie:** *Autori vyslovujú pod'akovanie Filipovi Balákovi za pomocné výpočty, ktoré prispeli k tvorbe predloženého príspevku.*

## Úvod

Na vyjadrenie súhrnu podmienok, v ktorých človek, prípadne celý národ žije, a ktoré zároveň spoluvytvára, sa používa termín *životná úroveň*. K tomuto termínu sa postupne vyvinuli rôzne definície, ktoré sú determinované predovšetkým disciplínou, v rámci ktorej daná definícia vznikla. Dokonca ani v jednotlivých disciplínach neexistuje v prístupe k životnej úrovni jednoznačná konzistencia. Hlavnými vednými odbormi, ktoré sa problematike životnej úrovne venujú, sú predovšetkým ekonómia, sociológia a psychológia.

Fidrmuc (1970) uvádza, že v tejto súvislosti je kľúčová analýza ľudských potrieb, ich charakteru, naliehavosti a hierarchie. Najznámejším modelom hierarchie potrieb je Maslowova pyramída, podľa ktorej prvou skupinou potrieb sú potreby fyziologické a poslednou potreby estetické. Hierarchia spočíva v tom, že najskôr musia byť uspokojené potreby najnižšie, potom potreby na vyššom stupni, avšak tie by bez uspokojenia základných potrieb ani nevznikli (Maslow, 1943). Nešporová, Svobodová a Vidovičová (2008) dodávajú, že základné potreby, nevyhnutné na prežitie, človek chápe ako samozrejmé, avšak kognitívne potreby, ku ktorým sa radí potreba učiť sa a poznávať, a estetické potreby, nebudú nikdy celkom naplnené, no túžba po nich stúpa, hoci boli sčasti uspokojené, a to preto, lebo prispievajú k rozvoju ľudského života a k jeho obohateniu.

Z *ekonomického* pohľadu životná úroveň tvoria položky, ktoré sú ovplyvňované merateľnými ukazovateľmi. Patria k nim najmä príjmy a výdavky, spotreba, kúpna sila, mzdy, nezamestnanosť, cenová hladina (Michaelson, Thomson a Marks, 2009). Túto kvantifikáciu a nadväznosť na matematické vedy vo vyjadrení životnej úrovne je vidieť aj vo vysvetlení Knausovej (2005), podľa ktorej sa *štatisticky* životná úroveň vyjadruje kvantifikovane, teda v merateľných veličinách. Charakterizujú ju predovšetkým príjmy, predpokladaná spotreba na osobu a podobne. Podstatné sú teda reálne mzdy, prípadne dôchodky obyvateľstva, čo sú faktory v neustálom dynamickom pohybe a meniace sa v čase.

Zo *sociologického* hľadiska je životná úroveň charakterizovaná ako stupeň uspokojovania životných potrieb obyvateľstva a ako súhrn všetkých podmienok, za ktorých sú tieto potreby uspokojované. Podstatnými ukazovateľmi životnej úrovne spoločnosti sú: úroveň výživy, bývania, odievania, dopravy, kultúry a rekreácie, vzdelania a kvalifikácie, zdravotnej a sociálnej starostlivosti, zamestnanosti a pracovných podmienok, spotreby a úspor, dodržiavania ľudských práv a podobne (Svobodová in Klimentová, 2009). Mnohí autori životnú úroveň spájajú s kvalitou života či blahobytom, alebo ju s nimi zamieňajú. V tejto súvislosti sa stretávame aj s pojmami *šťastie* a *spokojnosť*. Vzhľadom na stále nejasné vymedzenie týchto pojmov nie je však jednoduché presne určiť, akým spôsobom možno túto veličinu merať.

T a b u ľ k a 1

## Indexy používané na meranie životnej úrovne

Skupina indikátorov	Indikátor	Premenné
<i>Ekonomické</i>	Gross Domestic Product (GDP)	výdavky na konečnú spotrebu, tvorba hrubého kapitálu, čistý export
	<i>Genuine Saving Index</i> (GSI)	hrubé úspory, spotreba fixného kapitálu, čisté národné príjmy, výdavky na vzdelávanie, spotreba energií, spotreba minerálov, spotreba lesov, poškodenie emisiami, škody oxidu uhličitého
	<i>Global Competitiveness Index</i> (GCI)	inštitúcie, makroekonomické prostredie, zdravie a primárne vzdelávanie, efektivita trhu tovarov, efektivita trhu práce, rozvoj finančného trhu, technologická pripravenosť, veľkosť trhu, podnikateľská sofistikovanosť, inovácie
<i>Spoločenské</i>	<i>Human Development Index</i> (HDI)	očakávaná dĺžka života, priemerná dĺžka školskej dochádzky, očakávaná dĺžka školskej dochádzky, hrubý národný príjem
	<i>Quality of Life Index</i> (QLI)	kúpna sila, bezpečnosť, zdravotná starostlivosť, spotrebiteľské ceny, cena nehnuteľností voči príjmom, čas dochádzania do práce, znečistenie
	<i>Better Life Index</i> (BLI)	bývanie, príjem, práca, komunita, vzdelávanie, verejná angažovanosť, zdravie, spokojnosť, bezpečnosť, rovnováha pracovného a osobného života
	<i>Legatum Prosperity Index</i> (LPI)	hospodárstvo, sociálny kapitál, osobná sloboda, bezpečnosť, zdravie, vzdelávanie, vláda, podnikanie a príležitosti
	<i>Sustainable Society Index</i> (SSI)	stravovanie, vzdelávanie, zdravý život, rodová rovnosť, príjmová distribúcia, populačný rast, vláda, biodiverzita, obnoviteľné zdroje vody, kanalizácia, spotreba, využitie energií, úspora energií, skleníkové plyny, obnoviteľná energia, organické poľnohospodárstvo, čisté úspory, HDP, zamestnanosť, verejný dlh
	<i>Corruption Perceptions Index</i> (CPI)	AFDB, BF (SGI), BF (BTI), IMD, ICRG, WB, WEF, WJP, EIU, GI, PERC, TI, FH
<i>Environmentálne</i>	<i>Environmental Performance Index</i> (EPI)	zdravotné následky, kvalita ovzdušia, voda a kanalizácia, zdroje vody, poľnohospodárstvo, lesy, rybárstvo, biodiverzita a prírodné prostredie, klíma a energie
	<i>Happy Planet Index</i> (HPI)	očakávaná dĺžka života, blahobyt, ekologická stopa

AFDB – African Development Bank Governance Ratings; BF SGI – Bertelsmann Foundation Sustainable Governance Indicators; BFTI – Bertelsmann Foundation Transformation Index; EIU – Economist Intelligence Unit Country Risk Ratings; FH – Freedom House Nations in Transit; GI – Global Insight Country Risk Ratings; IMD – IMD World Competitiveness Yearbook; PERC – Political and Economic Risk Consultancy 2013; PRS – Political Risk Services International Country Risk Guide; TI – Transparency International Bribe Payers Survey 2011; WB – World Bank Country Policy and Institutional Assessment 2012; WEF – World Economic Forum Executive Opinion Survey (EOS) 2013; WJP – World Justice Project Rule of Law Index.

Prameň: Transparency International (2013).

Najpoužívanejším, a zároveň ostro kritizovaným ukazovateľom životnej úrovne je hrubý domáci produkt (HDP) prepočítaný na obyvateľa. Množstvo ekonómov považuje rast HDP na obyvateľa za najdôležitejší cieľ, ale Krugman a Wells (2012) tvrdia, že nie je dostatočný pri meraní ľudského blahobytu a sám osebe nie je ani vhodnou pomôckou pri politickom rozhodovaní. K tomu sa pripájajú

mnohí ďalší autori. Vznikajú rôzne alternatívne indexy, ktoré spresňujú či úplne nahrádzajú HDP na obyvateľa. Postupne sa ich vyvinulo veľké množstvo. Každý alternatívny index zohľadňuje rôzne množstvo determinantov, ktoré sa s niektorými v určitých častiach zhodujú, s inými vôbec nie. Pre veľké množstvo faktorov sme spracovali stručnejší prehľad vybraných indexov, ktoré sme uviedli v tabuľke 1.

Z dôvodu prehľadnosti sú podľa ich prevažného obsahu rozdelené do troch skupín – ekonomické, spoločenské a environmentálne.

Veľké množstvo faktorov vytvára multidimenzionalitu životnej úrovne a spôsobuje nejednoznačnosť jej vymedzenia. Tieto skutočnosti sa stali našou motiváciou – porovnať vybrané metódy určenia determinantov životnej úrovne a stanoviť výsledné kľúčové premenné, ktoré operacionalizujú pojem *životná úroveň* v podmienkach Európskej únie.

## Metodika

Hlavným zdrojom dát pre tento príspevok sú výsledky dotazníkového prieskumu z roku 2013, ktorým sa zisťovali subjektívne názory respondentov z vybraných piatich krajín EÚ (Česká republika, Fínsko, Francúzsko, Španielsko a Veľká Británia) na dôležitosť určených faktorov pre ich spokojnosť so životom. Na základe teoretických poznatkov z odborných štúdií bolo určených 99 faktorov životnej úrovne, ktoré sú zaradené do 8 oblastí – verejný život a záležitosti v rámci štátu, miesto bydliska, ekonomická stránka života, vzdelávanie, zdravotníctvo, životné prostredie, medziľudské vzťahy a osobnostné prvky. Ku každej položke sa respondenti mali vyjadriť, ako veľmi vplýva na ich životnú úroveň, a to pomocou desaťstupňovej škály, v ktorej 1 znamená najmenší vplyv a 10 najväčší vplyv, resp. dôležitosť. Respondenti sa vyjadrovali k spokojnosti s jednotlivými oblasťami, ako aj s celkovou životnou úrovňou pomocou Likertovej škály (veľmi spokojný – spokojný – skôr spokojný – skôr nespokojný – nespokojný – veľmi nespokojný). Zber primárnych dát, ktorý sa uskutočňoval pomocou elektronického systému Umbrela, vyvinutého špeciálne na tieto účely a v záujme dosiahnutia reprezentatívnej vzorky bol doplnený aj osobným prieskumom v teréne. Dotazník mala vyplniť vždy osoba na čele domácnosti.<sup>1</sup> Výber respondentov bol uskutočnený na základe kvót podľa ekonomickej aktivity – zamestnaný, samostatne zárobkovo činná osoba (SZČO), nezamestnaný, dôchodca, ostatné (Eurostat, 2014). Spolu sa výskumu zúčastnilo 2 743 respondentov. Tabuľka 2 približuje štruktúru výberového súboru.

<sup>1</sup> Touto osobou sa rozumie člen domácnosti, ktorého ostatní členovia označili za hlavného člena, prípadne člen, ktorý prispieva najväčším podielom k celkovému príjmu domácnosti.

T a b u ľ k a 2

**Štruktúra respondentov**

Identifikačné údaje	Absolútny počet	Relatívny počet (v %)
<b>Štát</b>	<b>2 743</b>	<b>100.00</b>
Česká republika	1 164	42.44
Fínsko	311	11.34
Francúzsko	473	17.24
Španielsko	584	21.29
Veľká Británia	211	7.69
<b>Ekonomická aktivita</b>	<b>2 743</b>	<b>100.00</b>
Zamestnaný	1 245	45.39
SZČO	469	17.10
Dôchodca	630	22.97
Nezamestnaný	178	6.49
Ostatné	221	8.06
<b>Najvyššie dosiahnuté vzdelanie</b>	<b>2 743</b>	<b>100.00</b>
Primárne	442	16.11
Sekundárne	1 215	44.29
Terciárne	1 086	39.59
<b>Pohlavie</b>	<b>2 743</b>	<b>100.00</b>
Žena	1 249	45.53
Muž	1 494	54.47

Prameň: Vlastné spracovanie.

Na stanovenie determinantov životnej úrovne využijeme analýzu hlavných komponentov (*Principal Component Analysis* – PCA), ktorá umožňuje zredukovať počet premenných bez veľkej straty pôvodných informácií. Hlavné komponenty sú lineárnymi kombináciami pôvodných premenných. Hlavné komponenty  $Z_j$  sa vytvárajú z  $p$  premenných  $X_i$ ; hlavné komponenty nie sú korelované a sú zoradené podľa veľkosti rozptylu.

Vhodnosť dát na analýzu hlavných komponentov posúdime pomocou Kaiseirovho-Meyerovho-Olkinovho indexu (KMO), ktorý vyjadruje mieru adekvátnosti výberu. Bertlettov test sféricity testuje, či nie sú premenné iba autokorelované a či majú aj ďalšie sféry vzťahov k ďalším premenným. V ďalšom kroku PCA určíme počet faktorov (hlavných komponentov) pomocou Cattellovho grafu vlastných čísel (sutinového diagramu). V diagrame je dôležitým miestom bod zlomu medzi kolmou a vodorovnou líniou vykreslenou z hlavných komponentov. Hodnota zlomu udáva počet významných komponentov.

Váhu premenných prisúdených každému faktoru určíme z faktorových záťaží a ich rotácie. Výsledná matica faktorových záťaží po rotácii faktorov obsahuje manifestné premenné, ktoré sú silne korelované iba s niektorými faktormi. Rotáciu je možné urobiť rôznymi spôsobmi, v tomto príspevku použijeme rotáciu *Varimax*, ktorá predstavuje ortogonálnu rotáciu eliminujúcu počet premenných, vyznačujúcich sa vysokými záťažami s každým spoločným faktorom.

Po uskutočnení predchádzajúcich krokov PCA dôjde k interpretácii faktorov, ktorá spočíva v tom, že manifestné premenné s vysokými faktorovými záťažami

sú si obsahovo blízke, na základe čoho budú formulované názvy pre konečné faktory životnej úrovne. Prostredníctvom populačného priemeru budú jednotlivým faktorom priradené váhy a z toho odvodená ich dôležitosť (Hendl, 2006; Meloun a Militký, 2012; Rimarčík, 2007).

Analýzou hlavných komponentov sa zredukuje počet premenných, ktoré majú najväčší vplyv na životnú úroveň. Na komparáciu prístupu sme použili taktiež klasifikačné metódy z danej oblasti, pričom sme využili úplný dátový súbor. V rámci tejto časti príspevku uskutočnime niekoľko testov s rôznymi klasifikačnými algoritmami a overíme vplyv redukcie premenných pomocou analýzy hlavných komponentov a pomocou metód *data mining* a *feature selection*.

Medzi vybrané algoritmy, ktoré budeme testovať, sme zaradili SVM (Support Vector Machine), KNN (K-nearest Neighbors Algorithm), Discriminant, C5.0 a Bayes. Snahou bude eliminovať tie atribúty, ktoré na posudzovaný výsledok (určenie životnej úrovne obyvateľstva) nemajú výrazný vplyv. Cieľom *feature selection* je teda znížiť zložitosť problému, prípadne zjednodušiť vytváraný matematický model redukciou vybraných parametrov.

## Metodický postup

Na analýzu 99 ukazovateľov hodnotených respondentmi v primárnom prieskume sme zvolili viacrozmernú štatistickú metódu. Analýza hlavných komponentov tak umožňuje pokryť taký rozsiahly koncept podmienený mnohými faktormi, akým je životná úroveň, efektívnym spôsobom. Zmysel tejto metódy spočíva v nahradení pôvodných premenných menším počtom latentných premenných, nazývaných aj *hlavné komponenty*, ktoré vysvetľujú rozptyl pôvodných premenných, pričom nie sú navzájom korelované.

Korelačná matica obsahuje hodnoty korelačných koeficientov v každej z dvojíc pozorovaných premenných. Položky so silnejšou koreláciou by mali tvoriť spoločný faktor. Na posúdenie vhodnosti použitia faktorovej analýzy pre dané dáta sme použili KMO mieru a Barlettov test sféricity. Nulovú hypotézu v Bartlettovom teste zamietame. To potvrdzuje aj hodnota KMO miery na úrovni 0,939, ktorá spĺňa podmienku, že je vyššia než 0,7. Možno teda skonštatovať, že analýza hlavných komponentov je pre tieto dáta vhodná.

V tabuľke 3 zobrazujúcej základné výsledky sú dostupné vlastné čísla, pričom každému vlastnému číslu prináleží jeden hlavný komponent. Na základe Kaiserovho kritéria je možné usúdiť, že počet faktorov, ktoré popisujú životnú úroveň, bol zredukovaný na 19. Tieto faktory popisujú celkom 68,5 % rozptylu.

Pri pohľade na kumulatívny rozptyl však vidíme, že po ôsmom komponente rastie pomalším tempom, a keďže cieľom analýzy je stanoviť v rámci možnosti čo najmenší počet premenných, výsledný počet premenných je možné obmedziť

na 8. Ich celková variabilita predstavuje takmer 53 %, čo je možné, vzhľadom na množstvo a charakter premenných, považovať za dostačujúce.

T a b u ľ k a 3

**Základné výsledky analýzy hlavných komponentov**

Vlastné čísla (faktory) extrakcia: hlavné komponenty				
komponenty	vlastné číslo	% celk. rozptylu	kumulativ. (vlast. číslo)	kumulativ. % rozptylu
1	26.162	26.426	26.162	26.426
2	6.856	6.925	33.017	33.351
3	4.988	5.039	38.006	38.390
4	3.770	3.808	41.775	42.197
5	3.251	3.284	45.026	45.481
6	2.634	2.661	47.661	48.142
7	2.464	2.489	50.124	50.631
8	2.246	2.268	52.370	52.899
9	1.854	1.873	54.224	54.772
10	1.801	1.819	56.026	56.592
11	1.643	1.660	57.669	58.251
12	1.585	1.601	59.254	59.852
13	1.479	1.494	60.733	61.346
14	1.361	1.375	62.094	62.721
15	1.226	1.238	63.319	63.959
16	1.217	1.229	64.536	65.188
17	1.122	1.133	65.659	66.322
18	1.099	1.110	66.757	67.432
19	1.069	1.080	67.827	68.512

*Prameň:* Vlastné spracovanie.

Faktorové záťažez naznačujú, ktoré premenné by mali tvoriť spoločný faktor. Na dosiahnutie lepšieho rozdelenia položiek sme použili rotáciu faktorov Variamax. Podľa tabuľky s faktorovými záťažami po rotácii je možné spojiť do jedného faktora vždy položky, ktorých faktorové záťažez sú väčšie ako 0,6. Samozrejme, vzhľadom na obmedzenie racionálneho uvažovania softvéru je možné vykonávať isté zmeny v priradení jednotlivých položiek na základe najvyššej faktorovej záťažez, ktorú pri položke nájdeme. V takomto prípade však uvedený krok nie je potrebný, a preto je možné skonštatovať, že bolo určených konečných 8 faktorov, ktoré obsahujú jednotlivé premenné približené v tabuľke 4. Počet faktorov sítě náhodne zodpovedá počtu vopred stanovených oblastí životnej úrovne, avšak ich obsah sa odlišuje.

V tejto súvislosti sa vynára potreba interpretácie určených komponentov. Premenné, ktoré sú obsiahnuté v prvom faktore, úzko súvisia so životným prostredím, čo sa týka jeho čistoty, ale aj dostupnosti a rozlohy prírodných častí. Zaujímavé je, že tento faktor takmer úplne kopíruje pôvodnú oblasť. Výnimku tvorí chýbajúca položka možnosti triedenia odpadu, ktorej faktorová záťažez mala hodnotu 0,5977; takže takmer splnila podmienku zaradenia danej položky do faktora. Vzhľadom na spomenuté premenné je tento faktor pomenovaný ako *faktor životného prostredia*.

T a b u ľ k a 4

## Váhy výsledných faktorov

č.	Názov faktora	Premenné
1	<i>Životné prostredie</i>	čistota ovzdušia, čistota vodných tokov, čistota prírody, čistota verejných priestorov, dostupnosť zelene a lesov, rozloha zelene v mieste bydliska, výskyt hluku v mieste bydliska
2	<i>Vzdelávacie aktivity</i>	dostupnosť verejných vzdelávacích zariadení, kvalita verejných vzdelávacích zariadení, využiteľnosť vyučovaných predmetov v praxi, úroveň učiteľov (odbornosť, morálne zásady, prístup), dostupnosť aktivít vo voľnom čase v oblasti vzdelávania, ich kvalita v oblasti vzdelávania, finančná náročnosť vzdelávania, možnosti zahraničných výjazdov v rámci štúdia, školský systém
3	<i>Verejný život</i>	činnosť vlády a ministerstiev, politická dôvera, korupcia, činnosť súdnictva, vymáhateľnosť práva, kvalita zákonov
4	<i>Obslužnosť obyvateľstva</i>	dostupnosť obchodov s iným tovarom než s potravinami, úroveň obchodov s iným tovarom, dostupnosť verejných služieb, úroveň verejných služieb, kvalita športových aktivít vo voľnom čase, dostupnosť kultúrnych aktivít vo voľnom čase, kvalita kultúrnych aktivít vo voľnom čase
5	<i>Zdravotníctvo</i>	dostupnosť zdravotníckych zariadení, kvalita zdravotníckych zariadení, komunikácia zdravotníckeho personálu, finančná náročnosť služieb v zdravotníctve, dostupnosť zdravotníckych pomôcok a liekov, ceny zdravotníckych pomôcok a liekov
6	<i>Komerčná sféra</i>	ceny výrobkov potravín a nápojov, ceny nepotravinárskych výrobkov, kvalita nepotravinárskych výrobkov, ceny služieb, kvalita služieb
7	<i>Nadštandard</i>	veľkosť dedičstva
8	<i>Emočná pohoda</i>	pocit všeobecnej spokojnosti, pocit istoty, pocit lásky, pocit uznania a úcty

Prameň: Vlastné spracovanie.

Druhý faktor obsahuje položky súvisiace so vzdelávaním, a to jednak z hľadiska dostupnosti inštitúcií, jednak ich kvality, ich finančnej náročnosti, ako aj vzdelávacích aktivít vo voľnom čase a medzinárodnej inštitucionálnej spolupráce. Oproti pôvodne stanovenej oblasti vypadli dve položky – dostupnosť a kvalita súkromných vzdelávacích zariadení. Tomuto faktoru je pridelený súhrnný názov *faktor vzdelávacích aktivít*.

V treťom faktore sa pôvodných 15 položiek zredukovalo na 6. Tento faktor zahŕňa také premenné, ako činnosť štátnej správy a súdnictva, politická dôvera, kvalita legislatívneho rámca, stav korupcie, a preto je označený ako *faktor kvality verejného života*.

Ďalší faktor predstavuje dostupnosť určitých obchodov a ich úroveň, ako aj dostupnosť a úroveň verejných služieb, ďalej možnosti a úroveň kultúrnych aktivít vo voľnom čase a kvalitu športových aktivít vo voľnom čase. Zaujímavé je, že pri kultúrnych aktivitách je zahrnutá ich dostupnosť aj ich kvalita, ale pri športových iba ich kvalita. Tento faktor je pomenovaný ako *faktor obslužnosti obyvateľov*.

Dostávame sa k piatemu faktoru, ktorý sa viaže na oblasť zdravotníctva – od dostupnosti a kvality zariadení vrátane ich personálu, cez finančnú náročnosť týchto služieb, až po dostupnosť a finančnú náročnosť zdravotníckych pomôcok a liekov. Tento komponent je nazvaný ako *faktor zdravotníctva*.



V poradí šiesty faktor je reprezentovaný cenami potravín a nápojov, ale aj cenami a kvalitou nepotravinárskych výrobkov a služieb. Zaujímavé je, že vypadla položka kvality potravín a nápojov, ale cena týchto produktov bola zahrnutá do výsledného faktora. Vzhľadom na skladbu položiek ide o *faktor komerčnej sféry*.

Siedmy komponent obsahuje jedinú položku, ktorou je veľkosť dedičstva. V prípade nižšieho nastavenia faktorovej záťaže by pod to spadala aj dostupnosť zahraničných produktov. Dedičstvo však môže životnú úroveň skutočne podstatne určovať. Dedičstvo významnej hodnoty má jednak vplyv po stránke ekonomickej či materiálnej, ale rovnako aj psychologickú. Keď niekto zdedí napríklad dom, môže sa zbaviť svojej hypotéky, a zároveň doceliť stav lepšej psychickej pohody. Tento faktor predstavuje aj akúsi zodpovednosť voči budúcim generáciám. Dá sa povedať, že hmotné náležitosti majú medzigeneračný vplyv na životnú úroveň. Vzhľadom na potrebu zovšeobecnenia je tento komponent pomenovaný ako *faktor nadštandardu*.

Posledný, ôsmy faktor je orientovaný na emočnú a osobnostnú stránku. Patrí sem pocit všeobecnej spokojnosti, pocit istoty, pocit lásky, pocit uznania a úcty. Vylúčená bola iba možnosť seberealizácie. Dostáva preto názov *faktor emočnej pohody*.

Každý z uvedených faktorov má však na životnú úroveň inú mieru vplyvu. V tomto prípade by išlo o 45 premenných zaradených do 8 faktorov. Vzhľadom na skutočnosť, že v prieskume sa zisťovala dôležitosť jednotlivých premenných, tento krok je založený na zistení aritmetického priemeru hodnotenia dôležitosti jednotlivých premenných a následne populačného priemeru položiek podľa jednotlivých faktorov. Ten potom predstavuje váhu jednotlivých faktorov. Výsledné hodnoty uvádzame v tabuľke 5.

T a b u ľ k a 5

**Váhy výsledných faktorov**

č.	Názov faktora	Populačný priemer (v abs. hodnote)	Váha faktoru (v %)
1	faktor životného prostredia (ŽP)	7.164523	13.30
2	faktor vzdelávacích aktivít (VA)	6.575688	12.21
3	faktor kvality verejného života (VŽ)	6.192855	11.50
4	faktor obslužnosti obyvateľov (OO)	6.491745	12.05
5	faktor zdravotníctva (Z)	7.510329	13.94
6	faktor komerčnej sféry (KS)	6.873861	12.76
7	faktor nadštandardu (N)	5.117390	9.50
8	faktor emočnej pohody (EP)	7.934378	14.73
<b>Spolu</b>	–	–	<b>100.00</b>

*Prameň:* Vlastné spracovanie.

Faktor emočnej pohody má najvyššiu váhu na úrovni 14,73 %. Po ňom nasleduje faktor zdravotníctva (13,94 %), faktor životného prostredia (13,30 %), faktor

komerčnej sféry (12,76 %), faktor vzdelávacích aktivít (12,21 %), ďalej faktor obslužnosti obyvateľov (12,05 %) a faktor kvality verejného života (11,50 %). Najnižšia váha s hodnotou 9,50 % je spojená s faktorom nadštandardu.

Pri hľadaní vhodnej metódy z oblasti *data mining* sme dáta testovali použitím algoritmov C5.0, Bayes Net, Discriminant, KNN a SVM z programu SPSS Modeler. Najskôr bola v programe zostavená sieť uzlov predstavujúcich jednotlivé kroky výpočtu presnosti klasifikácie testovaných algoritmov nad zvolenými dátami. Následne boli vytvorené a naučené modely využité v experimentoch.

Počas uskutočňovania experimentov bolo ponechané pri testovaných algoritmoch v SPSS Modeler pôvodné nastavenie pripravené programom, okrem algoritmov C5.0 a KNN. Pri nich bola nastavená metóda učenia pomocou krosvalidácie s rozdelením dátového súboru na 5 častí (Barrow a Crone, 2016).

Tabuľka 6

**Výsledky merania presnosti algoritmov**

Algoritmus	Presnosť (v %)
C5.0	99.74
Bayes Net	91.58
Discriminant	49.69
KNN	99.93
SVM	99.42

*Prameň:* Vlastné spracovanie.

Zo získaných výsledkov uvedených v tabuľke 6 je zrejme, že najpresnejšou metódou pre dané dáta je algoritmus KNN s presnosťou 99,93 %. Spolu s týmto algoritmom sú tiež veľmi dobre využiteľné algoritmy C5.0 a SVM, ktoré dosiahli presnosť klasifikácie vyššiu než 99 %.

Výsledné premenné sa pri použití *data mining* a analýze hlavných komponentov nezhodujú. Hlavným rozdielom je konečný počet premenných, ktorý je pri PCA 45. V prípade *data mining* bolo vylúčených 30 atribútov, a teda konečný počet faktorov, ktoré ovplyvňujú životnú úroveň, je 69. Tabuľka 7 približuje rozdiely vo vylúčených premenných pri použití týchto metód. Niektoré premenné boli vylúčené obomi metódami, navyše boli vylúčené odlišné faktory v prípade PCA a *data mining*.

Na redukciu atribútov pomocou *data mining* bola zvolená metóda *feature selection* založená na teste False discovery rate. Pomocou skriptu implementovaného v Pythone pomocou knižnice Scikit-learn bolo odobratých celkom 30 atribútov z pôvodného dátového súboru. Táto implementácia výberu atribútov sa osvedčila viac než nástroj *feature selection* v SPSS Modeler, keď bolo vyradených iba 7 atribútov.

T a b u ľ k a 7

**Zoznam vylúčených premenných podľa použitej metódy**

Premenné vylúčené oboma metódami	Dodatočné premenné vylúčené AHK (PCA)	Dodatočné premenné vylúčené podľa data mining
kvalita dopravnej obslužnosti, ceny zahraničných výrobkov, činnosť polície, dostupnosť domovov pre seniorov, domovov pre osoby so zdrav. postihnutím, dostupnosť dopravnej obslužnosti v rámci štátu (napr. vlaky, autobusy medzi obcami), dostupnosť energií v mieste bydliska, dostupnosť internetu na verejných miestach, dostupnosť zahraničných výrobkov, kvalita ciest a chodníkov, kvalita domovov pre seniorov a domovov pre osoby so zdrav. postihnutím, kvalita dopravnej obslužnosti v rámci štátu, kvalita výrobkov potravín a nápojov, množstvo pracovných príležitostí, možnosti bývania ponúkané štátom a obcami, možnosti triedenia odpadu, vlastníctvo auta, vlastníctvo telefónu a počítača, výška daňových odvodov	množstvo cudzincov a menších v štáte, dostupnosť medzinárodnej dopravy, kvalita medzinárodnej dopravy, kvalita pozemných komunikácií (diaľnice, cesty medzi okresmi a obcami), aktivity neziskových organizácií, činnosť starostu a zastupiteľstva, činnosť miestnych úradov, správanie úradníkov voči občanom, modernizácia a rozširovanie výstavby, bezpečnosť, výskyt bezdomovcov, úroveň obchodov s potravinami, dostupnosť športových aktivít vo voľnom čase, cena bývania (nájomné, poplatky za energie a ostatné poplatky spojené s bývaním), výška soc. transferov pre domácnosť (dávky a finančná podpora od štátu a obce), veľkosť priemerných úspor domácnosti, vlastníctvo nehnuteľností, rovnováha medzi prácou a voľným časom, dostupnosť a kvalita súkromných vzdelávacích zariadení, vzťahy v rodine, vzťahy na pracovisku, vzťahy v susedstve, zdravie v rodine a medzi priateľmi, vplyv dedičstva na vzťahy v rodine, etika v spoločnosti (slušné správanie detí, zamestnancov v obchodoch a pod.), usporiadanie hodnôt v spoločnosti, vzťah, schopnosti, znalosti, talent, postavenie v spoločenskom rebríčku, najvyššie dosiahnuté vzdelanie, možnosť sebarealizácie	ceny nepotravinárskych výrobkov (napr. oblečenie), činnosť vlády a ministerstiev, čistota ovzdušia, čistota prírody (parky, lesy, lúky), čistota vodných tokov, dostupnosť zdrav. pomôcok a liekov, finančná náročnosť vzdelávania, kvalita zákonov, vymožiteľnosť práva, výskyt hluku v mieste bydliska, úroveň obchodov s iným tovarom než potravinami

*Prameň:* Vlastné spracovanie.

T a b u ľ k a 8

**Výsledky merania presnosti klasifikačných algoritmov s využitím výberu atribútov**

Algoritmus	Presnosť (v %)
C5.0	99.89
Bayes Net	86.84
Discriminant	46.45
KNN	99.89
SVM	96.39

*Prameň:* Vlastné spracovanie.

Z výsledkov merania presnosti algoritmov po aplikovaní metód výberu atribútov (pozri tab. 9) sme zistili, že jediný algoritmus, ktorému sa zlepšila presnosť, je C5.0. Ostatným sa presnosť znížila.

Napriek tomu, že algoritmus KNN bez výberu atribútov má o niekoľko desiatín lepšiu presnosť, na klasifikáciu ďalších odpovedí z dotazníku je možné využiť aj algoritmus C5.0. Po výbere atribútov dosahuje rovnako vysokú presnosť. Pri oboch algoritmoch po aplikovaní výberu atribútov by malo byť znížené riziko pretrénovania.

T a b u ľ k a 9

**Výsledky merania presnosti klasifikačných algoritmov s využitím analýzy hlavných komponentov**

Algoritmus	Presnosť (v %)
C5.0	99.96
Bayes Net	82.36
Discriminant	35.76
KNN	99.89
SVM	84.69

*Prameň:* Vlastné spracovanie.

Po redukcii atribútov pomocou analýzy hlavných komponentov sa však klasifikátorom okrem C5.0 znížila presnosť klasifikácie. To bolo pravdepodobne spôsobené veľmi veľkým množstvom vyradených atribútov z pôvodného súboru. Paradoxne však úspešnosť klasifikácie pomocou C5.0 vzrástla na 99,96 %. Pri takejto vysokej presnosti sa ponúka otázka, či naučený model nie je pretrénovaný.

**Záver**

Spracovanie výsledkov primárneho výskumu pomocou metódy hlavných komponentov umožnilo určiť 8 výsledných faktorov, ktoré sú určujúce a majú preukázateľný vplyv na životnú úroveň obyvateľov. Týchto 8 faktorov pokrýva celkom 45 konečných premenných, čo znamená, že bolo vylúčených 54 premenných. Došlo teda k výraznému zjednodušeniu modelu životnej úrovne, čo ale stále znamená, že na posúdenie jej úrovne je potrebné vyhodnocovať 45 premenných. Na tomto mieste sa ponúka otázka, akým spôsobom je možné tieto premenné kvantifikovať. Existujú rôzne možnosti, ktoré sú používané už v iných alternatívnych prístupoch k posudzovaniu životnej úrovne, ale aj tie sa od seba odlišujú. Ide teda o podnet na ďalšie skúmanie v tejto oblasti.

Z uskutočnených analýz sme zistili, že najväčšiu váhu má faktor emočnej pohody. V prípade, že by sa v dotazníkovom prieskume zisťovala spokojnosť s danými položkami, bolo by možné zostrojiť regresný model, kde by závislou premennou bola spokojnosť a nezávislými premennými, spokojnosť s konkrétnou položkou. V tomto príspevku však nebolo toto zistenie prioritné.

Zistené skutočnosti sa približujú k výsledkom viacrozmerného chápania blahobytu v podmienkach EÚ podľa Stiglitzovej-Senovej-Fitoussiho správy (Stiglitz, Sen a Fitoussi, 2009), podľa ktorej je životná úroveň determinovaná aj inými než len ekonomickými vonkajšími faktormi. Za zhodné je považované životné prostredie, zdravie, vzdelanie, politické prostredie a kvalita verejnej správy (verejný život a obsluha), a sociálne kontakty a vzťahy (emočná pohoda a osobná aktivita). Pri zohľadnení alternatívnych indikátorov, ktoré určitým spôsobom

upravujú, dopĺňajú alebo úplne nahrádzajú HDP na obyvateľa ako ukazovateľ životnej úrovne, si možno všimnúť, že výsledky sú si najviac podobné s faktormi zahrnutými v Better life indexe vyvinutom OECD.

Životná úroveň je koncept, ktorého determinanty sú skutočne rôznorodé a pri odhaľovaní všetkých spojitostí je potrebné ísť do hĺbky a skúmať mnohé premenné. Našou snahou bolo zjednodušiť model tak, aby boli pokryté všetky významné oblasti.

Metódu *data mining* sme využili hlavne na kontrolu vybraných faktorov metódou hlavných komponentov. Skúmali sme, ako redukcia faktorov ovplyvní schopnosť priradiť určitú triedu (spokojnosť so životnou úrovňou) na základe kombinácie odpovedí v dotazníku. Prostredníctvom tohto prístupu bolo vylúčených 30 premenných, a teda konečný stanovený počet je 69 premenných ovplyvňujúcich životnú úroveň.

Je zrejmé, že medzi analýzou hlavných komponentov a *feature selection* existujú určité rozdiely. Analýza hlavných komponentov vyraduje viac atribútov a znižuje čas určenia pri klasifikácii. Táto metóda nezohľadňuje hodnoty cieľového atribútu (učenia bez učiteľa). Oproti tomu *feature selection* vyradí menej atribútov, berie do úvahy hodnoty cieľového atribútu a umožňuje postup zautomatizovať.

Rozdielnosť použitých metód, vzhľadom na redukciiu posudzovaných atribútov, je daná nastavením čiastkových parametrov konkrétnej metódy. Vplyv má hlavne použitá metrika vzdialenosti danej metódy, ktorá bezprostredne ovplyvňuje tvorbu zhlukov. Táto skutočnosť má za následok veľkosť realizovanej redukcie (zníženie počtu atribútov). Kvalitu redukcie atribútov, t. j. odstránenie nadbytočných atribútov, so zreteľom na stálosť klasifikačnej metódy, je možné overiť jednoznačne tak, že výsledok klasifikácie sa nezmení ani pre pôvodný počet atribútov, a následne ani pre ich redukovaný variant.

Podľa výsledkov sa ako výhodnejšie javí použitie analýzy hlavných komponentov. Presnosť niektorých algoritmov síce klesne, ale algoritmy, ktoré využili krosvalidáciu, mali presnosť rovnakú, alebo vyššiu než pri použití *feature selection* (KNN a C5.0).

Porovnaním vylúčených premenných prostredníctvom analýzy hlavných komponentov a *data mining* sme dospeli k zisteniu, že redukcia premenných ako nevýznamných na vyjadrenie životnej úrovne obyvateľov sa zhoduje len pri 19 zo sledovaných 99 premenných. Ide o premenné skutočne nevýznamné (či považované za samozrejmé), ako sú vlastníctvo auta, vlastníctvo telefónu, kvalita ciest a chodníkov, cena a dostupnosť zahraničných produktov a ďalšie. Potom sú to premenné, pri ktorých obyvatelia rezignujú na dôležitosť, napr. činnosť polície, kvalita dopravnej obsluhy, množstvo pracovných príležitostí, výška

daňových odvodov a iné. Odpoveďou na spochybnenie jednoznačnosti určenia významu faktorov životnej úrovne by bola korelácia významu faktora a spokojnosti s ním.

Z ôsmich významovo blízkych komponentov – faktorov bol iba jeden, pri ktorom nebola žiadna premenná spochybnená ani jednou z použitých metód. Ide o faktor emočnej pohody.

K identifikácii determinantov životnej úrovne možno dospieť iba v spojení dôležitosti premenných s vyjadrením s ich spokojnosťou. Použitie viacerých metód na kvantifikáciu životnej úrovne má funkciu skôr kontrolnú či podpornú než funkciu konfrontačnú.

Z výsledkov je zrejmé, že životná úroveň nie je iba ekonomickou problematikou, ale zároveň sa od ekonomického prístupu nemožno úplne odkloniť. Napriek tomu, že ekonomické položky obyvatelia vybraných štátov nehodnotili ako najdôležitejšie, tvoria základ pre život každého jedinca. Ekonomická dimenzia je stále nosnou časťou pri posudzovaní životnej úrovne. V sledovaných krajinách, oproti rozvojovým častiam sveta, ľudia kriticky neriešia problematiku uspokojovania základných fyziologických potrieb, a teda dochádza k uspokojovaniu aj menej dôležitých potrieb, čo je hlavným dynamickým princípom Maslowovej pyramídy. Významné miesto pri vysvetlení týchto zistení zohrávajú spoločenské konvencie spojené so sociálnou komparáciou, ktorá vysvetľuje Easterlinov paradox (Easterlin, 1995). Znamená to, že rast príjmov nevyvoláva rast spokojnosti ľudí, ktorí porovnávajú svoj blahobyt s blahobytom spoločnosti ako celku.

## Literatúra

- AHN, H. (2014): Optimization of Multiclass Support Vector Machine Using Genetic Algorithm: Application to the Prediction of Corporate Credit Rating. *Information Systems Review*, 16, č. 3, s. 161 – 177.
- AL SHAQSI, J. – WANG, W. J. (2013): Estimating the Predominant Number of Clusters in a Dataset. *Journal Intelligent Data Analysis*, 17, č. 4, s. 603 – 626.
- BARROW, D. K. – CRONE, S. F. (2016): Cross-validation Aggregation for Combining Autoregressive Neural Network Forecasts. *International Journal of Forecasting*, 32, č. 4, s. 1120 – 1137.
- DIJKSTRA, T. (2014): Ridge Regression and Its Degrees of Freedom. *Quality & Quantity*, 48, č. 6, s. 3185 – 3193.
- EASTERLIN, R. A. (1995): Will Raising the Incomes of All Increase the Happiness of All. *Journal of Economic Behavior & Organization*, 27, č. 1, s. 35 – 47.
- EUROSTAT (2014): European Statistics. Dostupné z: <http://ec.europa.eu/eurostat/data/database>.
- FIDRMUC, J. (1970): Blahobyt a ekonomický rast. Bratislava: Slovenské pedagogické nakladateľstvo.
- HENDL, J. (2006): Přehled statistických metod zpracování dat. Praha: Portál. ISBN 80-7367-123-9583.
- KLIMENTOVÁ, E. (2009): Sociální politika I. Olomouc: Univerzita Palackého v Olomouci.
- KNAUSOVÁ, I. (2005): Vybrané kapitoly ze sociální politiky. 2 vyd. Olomouc: Univerzita Palackého v Olomouci. ISBN 80-244-1021-4.

- KRUGMAN, P. R. – WELLS, R. (2012): *Macroeconomics*. 3. vyd. New York: Worth Publishers. ISBN 978-1-4292-8343-4.
- LUO, Y. – LIU, T. L. – TAO, D. C. – XU, C. (2015): Multiview Matrix Completion for Multilabel Image Classification. *IEEE Transactions on Image Processing*, 24, č. 8, s. 2355 – 2368.
- MASLOW, A. H. (1943): A Theory of Human Motivation. *Psychological Review*, 50, č. 4, s. 370 – 396.
- MELOUN, M. – MILITKÝ, J. (2012): *Kompendium statistického zpracování dat*. 3. vyd. Praha: Karolinum. ISBN 978-80-246-2196-8.
- MICHAELSON, A. – THOMSON, S. – MARKS, N. (2009): *National Accounts of Well-being: Bringing Real Wealth onto the Balance Sheet*. London: New Economics Foundation. ISBN 978-1-904882-50-3.
- NEŠPOROVÁ, O. – SVOBODOVÁ, K. – VIDOVIČOVÁ, L. (2008): *Zajištění potřeb seniorů s důrazem na roli nestátního sektoru*. Praha: Výzkumný ústav práce a sociálních věcí, v.v.i. 85 s. ISBN 978-80-87007-96-9.
- ORANTA, O. – ROUTASALO, P. – HUPLI, M. (2002): Barriers to and Facilitators of Research Utilization Among Finnish Registered Nurses. *Journal of Clinical Nursing*, 11, č. 2, s. 205 – 213.
- RIMARČÍK, M. (2007): *Štatistika pre prax*. Vydáno vlastním nákladem. ISBN 8096981311.
- TIRIVANHU, P. – MATONDI, P. B. – GROENEWALD, I. (2015): Comprehensive Community Initiative: Evaluation of a Transformation System in Mhakwe Community in Zimbabwe. *Development Southern Africa*, 32, č. 6, s. 785 – 800.
- TRANSPARENCY INTERNATIONAL (2013): *Corruption Perceptions Index: Full Source Description*. [Online.] Dostupné z: <[https://www.transparency.org/files/content/pressrelease/2013\\_CPISourceDescription\\_EN.pdf](https://www.transparency.org/files/content/pressrelease/2013_CPISourceDescription_EN.pdf)>.
- STIGLITZ, J. E. – SEN, A. – FITOUSSI, J. P. (2009): *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Dostupné na: <<http://www.stiglitz-sen-fitoussi.fr/en/documents.htm>>.
- ZHANG, Y. – GONG, D. – ZHANG, W. (2016): Feature Selection of Unreliable Data Using an Improved Multi-objective PSO Algorithm. *Neurocomputing*, 171, Issue C, s. 1281 – 1290.