

SLOVENSKÁ ŠTATISTIKA a DEMOGRAFIA

SLOVAK STATISTICS
and DEMOGRAPHY

3/2020

ročník/volume 30

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: 9 – 39

Dátum vydania/Publication date: 15. júl 2020/July 15, 2020



Iveta STANKOVIČOVÁ
Fakulta managementu, Univerzita Komenského v Bratislave
Alena MOJSEJOVÁ
Ekonomická fakulta, Technická univerzita v Košiciach

INDEX ĽUDSKÉHO ROZVOJA: HODNOTENIE A KLASIFIKÁCIA EURÓPSKYCH KRAJÍN POMOCOU VIACROZMERNÝCH METÓD

HUMAN DEVELOPMENT INDEX: RATING AND CLASSIFICATION OF EUROPEAN COUNTRIES BY MULTIVARIATE METHODS

ABSTRAKT

Kvalita života a ľudský rozvoj v krajinách sveta sú v poslednej dobe často diskutované témy. Ide o zložené javy, ktoré vieme merať len pomocou viacerých ukazovateľov. Experti z Rozvojového programu Organizácie Spojených národov (UNDP) navrhli metodiku výpočtu integrálneho ukazovateľa na hodnotenie krajín sveta v oblasti ľudského rozvoja. Ukazovateľ sa volá index ľudského rozvoja (HDI) a v súčasnosti sa skladá zo štyroch čiastkových ukazovateľov, ktoré vieme zmerať. Na základe dosiahnutých hodnôt HDI, navrhli aj klasifikáciu krajín (resp. regiónov) do skupín. Cieľom príspevku je porovnať úroveň HDI v 34 krajinách Európy a vytvoriť hodnotiaci rebríček. Na tento účel použijeme hodnoty HDI vypočítané podľa metodiky UNDP, ale aj HDI vypočítaný viacrozmernými štatistickými metódami, ako sú analýza hlavných komponentov a faktorová analýza. Na klasifikáciu krajín do skupín použijeme zhlukovú analýzu. Výsledky porovnáme a zhodnotíme výhody a nevýhody týchto rôznych prístupov pri hodnotení ľudského rozvoja.

ABSTRACT

Quality of life and human development in the countries of the world have recently been frequently discussed. This compound phenomena can be measured only by using several indicators. Experts from the United Nations Development Program (UNDP) have proposed a methodology for calculating an integral indicator for assessing the countries of the world in the area of human development. The indicator is called the Human Development Index (HDI) and currently consists of four measurable sub-indicators. Based on the achieved HDI values, they also proposed the classification of countries (or regions) into groups. The aim of the paper is to compare the HDI values in 34 European countries and to create a ranking. For this purpose, we will use HDI values calculated by the UNDP methodology, but also the HDI calculated using multivariate statistical methods such as principal component analysis and factor analysis. We will use cluster analysis for the classification of countries into groups. We will compare the results and evaluate the advantages and disadvantages of these different approaches for evaluating human development.

KLÚČOVÉ SLOVÁ

index ľudského rozvoja, analýza hlavných komponentov, faktorová analýza, zhluková analýza, SAS Enterprise Guide

KEY WORDS

human development index, principal component analysis, factor analysis, cluster analysis, SAS Enterprise Guide

1. ÚVOD

Súčasná doba je charakteristická komparáciami rôznych javov. Často sa navzájom porovnávajú ekonomiky krajín, resp. zoskupení. Ako však môžeme jednoducho porovnávať národné ekonomiky? V minulosti sa používali hlavne makroekonomické ukazovatele, ako napríklad hrubý domáci produkt, inflácia, miera nezamestnanosti alebo zamestnanosti, platobná bilancia a dlh (pozri [4, 7, 10, 13]). Problémom týchto ukazovateľov je ich zameranie na stav a vývoj hospodárstva v krajine, ako na stav a vývoj krajiny ako celku. Makroekonomické premenné nehovoria o kvalite života a ľudskom rozvoji občanov v sledovanej krajine (pozri [5]).

Ľudský rozvoj je predovšetkým o snahe umožniť ľuďom viesť životy, ktoré si cenia a ktoré im umožňujú uvedomiť si ich potenciál ako ľudských bytostí. Trvalo udržateľné rozvojové ciele sú ciele v medzinárodne dohodnutom súbore čiastkových cieľov (známy pod názvom Agenda 2030, [15]) na zníženie extrémnej chudoby a rozšírenie rodovej rovnosti, rovnosti príležitostí na zdravie a vzdelávanie. Redukovanie nerovností v rôznych oblastiach sa stalo kľúčovým cieľom v Agende 2030.

Porovnávanie krajín na základe rôznych ukazovateľov predstavuje dôležitý pohľad na postavenie jednotlivých štátov a analýzu nerovnosti medzi nimi. Nerovnosť medzi krajinami ale aj v ich vnútri krajín je naďalej vážnym problémom napriek pokroku a snahám o znižovanie týchto rozdielov.

Nerovnosti v oblasti ľudského rozvoja sú najviac viditeľné. Podľa Správy o ľudskom rozvoji [18], nerovnosti v tejto oblasti poškodzujú spoločnosť, oslabujú sociálnu súdržnosť, dôveru ľudí vo vládu a v inštitúcie, ale aj vzájomnú dôveru ľudí. Navyše, tieto nerovnosti sú prekážkou pri dosahovaní cieľov Agendy 2030 pre trvalo udržateľný rozvoj.

Cieľom príspevku je analyzovať stav a vývoj ľudského rozvoja v európskych krajinách na základe známej a použíwanej miery, konkrétne indexu ľudského rozvoja (HDI). Téma analýzy ľudského rozvoja je veľmi aktuálna, o čom svedčí aj množstvo publikácií, ktoré sa danej tematike venujú, napr. [1, 2, 3]. V prácach [3] a [18] autori pomocou zhlukovej analýzy klasifikujú krajiny sveta do skupín podľa dosiahnutého stupňa ľudského rozvoja meraného pomocou HDI. Zhlukovej analýze európskych krajín podľa HDI na regionálnej úrovni sa venuje práca [9].

2. MERANIE ĽUDSKÉHO ROZVOJA

Ľudský rozvoj a kvalita života sú veľmi blízke pojmy a veľmi často sa s nimi stretávame v spoločných analýzach. Kvalita života je výsledkom vzájomného pôsobenia sociálnych, zdravotných a environmentálnych podmienok, týkajúcich sa ľudského a spoločenského rozvoja. Pojem kvalita života sa prvýkrát objavil v práci amerického ekonóma J. K. Galbraitha, ktorému sa pripisuje jeho autorstvo.

Dosiahnuté výsledky v oblasti ľudského rozvoja vo vyše 150 krajinách sveta, od roku 1990 monitoruje Rozvojový program Organizácie Spojených národov (UNDP – United Nations Development Programme). Ľudský rozvoj je podľa Organizácie spojených národov (OSN) definovaný ako proces rozširovania ľudských možností a obsiahnutý stupeň blahobytu. Pri ľudskom rozvoji rozlišujeme dva pohľady na získané schopnosti. Na jednej strane je to formovanie nadania, ktoré sa môže merať kvalitným vzdelávaním alebo dostupnou zdravotnou starostlivosťou. Na druhej strane uvažujeme využitie získaných schopností v pracovnom i voľnom čase (definícia podľa UNDP). Tieto možnosti sa môžu neustále meniť a pribúdať alebo sa môžu stať neaktuálnymi. V tomto prípade sa jedná o tri nevyhnutné činitele (podľa [17 a 19]):

- dĺžka života obyvateľstva danej krajiny,
- dosiahnutá vzdelanostná úroveň obyvateľstva,
- kvalita života ľudí vyjadrená reálnym HDP pripadajúcim na jedného obyvateľa.

2.1 MERANIE ĽUDSKÉHO ROZVOJA PODĽA METODIKY UNDP

Index ľudského rozvoja (z angl. Human Development Index – HDI) môžeme zaradiť medzi komplexné indikátory trvalo udržateľného rozvoja sociálno-ekonomickej povahy. Prvýkrát bol publikovaný v roku 1990 v správe Human Development Report. Hodnotí dosiahnuté priemerné výsledky krajiny v troch dimenziách ľudského rozvoja: dlhý a zdravý život (index zdravia – IH), prístup k vedomostiam (index vzdelania – IE) a primeraný životný štandard (príjmový index – II). Zhrnutie jednotlivých indexov do jedného ukazovateľa sa vytvára na báze kompozitného indikátora, pričom UNDP používa rovnaké váhy pre všetky tri ukazovatele (indexy). Výsledný HDI sa vypočítava ako geometrický priemer troch indexov, tzv. dimenzií:

$$HDI = \sqrt[3]{Index\ zdravia * Index\ vzdelania * Príjmový\ index} . \quad (1)$$

Po viacerých diskusiách odborníkov bola zvolená *očakávaná dĺžka života pri narodení* (angl. life expectancy at birth) ako najvhodnejší indikátor prvej dimenzie predstavujúcej "dlhý a zdravý život". Má dôležitú výpovednú hodnotu, keďže vek obyvateľstva krajiny môže vyjadrovať v určitých prípadoch napríklad aj zdravotný status ľudí tu žijúcich. Druhá dimenzia predstavujúca vzdelanie je vyjadrená mierou gramotnosti. Aktuálnymi indikátormi sú dva ukazovatele: *predpokladané roky štúdia* (angl. expected years of schooling) a *priemerný počet rokov vzdelávania* (angl. mean years of schooling). Poslednú dimenziu predstavuje kúpyschopnosť obyvateľstva, čiže uspokojovanie základných potrieb. Ukazovateľom, ktorý sa využíva v terajšom vzťahu je hrubý národný produkt (HNP) na obyvateľa vyjadrený v PPP \$¹ (angl. GNI per capita in PPP \$). Tento ukazovateľ nahradil predtým používaný logaritmus hrubého domáceho produktu (HDP).

¹ *PKS – parita kúpnej sily (angl. PPP – Purchasing Power Parity) je ekonomický ukazovateľ, na základe ktorého je možné porovnávať ekonomickú silu jednej krajiny (regiónu) oproti inej. Je to špecifický prevodový cenový index, ktorý dáva do súvisu cenové rozdiely tovarov a služieb v určitom okamihu v rôznych krajinách (regiónoch) pri eliminovaní rozdielov v cenovej úrovni medzi krajinami. V praxi sa často využíva ukazovateľ HDP na obyvateľa v parite kúpnej sily (HDP/PPS alebo angl. GDP/PPP), vyjadrený v US dolároch v parite kúpnej sily.*

Podľa viacerých autorov (pozri napr. [1, 6, 8, 13]), HDI poskytuje presný spôsob výpočtu ľudského rozvoja, čo je rozhodujúce pri meraní celkovej výkonnosti ktorejkoľvek krajiny. HDI navyše umožňuje ľahké a spravodlivé hodnotenie krajín na základe získanej hodnoty.

Výpočet výsledného indexu ľudského rozvoja HDI možno podľa [19] zhrnúť do dvoch krokov. V tabuľke č. 1 je zoznam ukazovateľov, ktoré sa používali v minulosti a ktoré sa používajú v súčasnosti (od roku 2009). Z týchto štyroch ukazovateľov sa počíta výsledná hodnota HDI.

Tabuľka č. 1: Porovnanie komponentov indexu ľudského rozvoja (HDI)

Pôvodný index z roku 1990		Index aktuálne využívaný (od roku 2009)		Index
Komponent	Indikátor	Dimenzia	Indikátor	
Dlhý život	Očakávaná dĺžka života pri narodení	Dlhý a zdravý život	Očakávaná dĺžka života pri narodení	Index zdravia
Vedomosti	Miera gramotnosti	Vedomosti	Očakávaný počet rokov štúdia Priemerný počet rokov štúdia	Index vzdelania
Kúpna sila	Logaritmus hrubého domáceho produktu na obyvateľa	Primerané životné štandardy	Hrubý národný produkt na obyvateľa v PPP \$ (2011)	Príjmový index

Zdroj: vlastné spracovanie podľa [4] a [6]

Prvým krokom je vytvorenie tzv. indexu dimenzie. Pre každý zo štyroch indikátorov odborníci určili hranice, minimálnu a maximálnu hodnotu (tabuľka č. 2).

Tabuľka č. 2: Minimálne a maximálne hodnoty komponentov HDI

Dimenzia	Indikátor (ukazovateľ)	Min.	Max.
Zdravie	Očakávaná dĺžka života pri narodení (v rokoch)	20	85
Vzdelanie	Očakávaný počet rokov štúdia	0	18
	Priemerný počet rokov štúdia	0	15
Primerané životné náklady	Hrubý národný produkt na obyvateľa (2011 PPP \$)	100	75 000

Zdroj: spracované podľa [19]

Podľa [19] tieto hranice sú určené kvôli preškálovaniu indikátorov vyjadrených v rôznych jednotkách na škálu od 0 po 1. Minimálna hodnota očakávanej dĺžky života je stanovená na 20 rokov, keďže žiadna krajina v 20. storočí nedosahovala nižšiu hodnotu. Maximálna dĺžka života je stanovená na 85 rokov, čo predstavuje cieľ pre mnoho krajín za posledných 30 rokov. Spoločnosti môžu existovať bez formálneho vzdelania, čo odôvodňuje nastavenie minimálnej hodnoty očakávaného počtu rokov štúdia na hodnotu 0. Maximálne očakávaný počet rokov štúdia je 18, čo je rovnocenné získaniu druhého stupňa vysokoškolského štúdia vo väčšine krajín. Maximálna hodnota priemerného počtu rokov štúdia je 15, čo je predpokladané maximum tohto ukazovateľa na rok 2025. Nízka minimálna hodnota hrubého národného dôchodku na obyvateľa je 100 dolárov. Podľa UNDP predstavuje nezmerateľnú časť HND, ktorá nie je zachytená v oficiálnych štatistikách. Maximálna výška je stanovená na 75 000 dolárov na obyvateľa. V súčasnosti (r. 2018) iba štyri krajiny (Brunej, Lichtenštajnsko, Katar a Singapur)

presahujú HND vo výške 75 000 dolárov na obyvateľa. Je však dokázané, že príjem nad 75 000 dolárov neprináša vysoký zisk pre ľudský rozvoj v krajine.²

Po definovaní minimálnej a maximálnej hodnoty, môžeme každý indikátor preškalovať na hodnoty z intervalu <0, 1> podľa nasledujúceho vzorca:

$$\text{Index dimenzie} = \frac{\text{aktuálna hodnota} - \text{minimálna hodnota}}{\text{maximálna hodnota} - \text{minimálna hodnota}} \quad (2)$$

V prípade indexu vzdelania, ktorý sa skladá až z dvoch čiastkových ukazovateľov (indikátorov), sa preškalovaný index vypočíta pre každý indikátor zvlášť a výsledný index je priemerom týchto dvoch hodnôt.

Výpočty čiastkových indexov a HDI ilustrujeme na príklade Slovenska (výpočty na rok 2018):

$$\text{Index zdravia} = \frac{77,4 - 20}{85 - 20} = 0,883 \quad (3)$$

$$\text{Index očakávaného počtu rokov štúdia} = \frac{14,5 - 0}{18 - 0} = 0,806 \quad (4)$$

$$\text{Index priemerného počtu rokov štúdia} = \frac{12,6 - 0}{15 - 0} = 0,840 \quad (5)$$

$$\text{Index vzdelania} = \frac{0,806 + 0,840}{2} = 0,823 \quad (6)$$

$$\text{Príjmový index} = \frac{\ln(30672) - \ln(100)}{\ln(75000) - \ln(100)} = 0,865 \quad (7)$$

Tabuľka č. 3: Vstupné hodnoty ukazovateľov na výpočet komponentov HDI pre Slovensko (rok 2018)

Indikátor	Hodnota
Očakávaná dĺžka života pri narodení (v rokoch)	77,4
Očakávaný počet rokov štúdia	14,5
Priemerný počet rokov štúdia	12,6
Hrubý národný produkt na obyvateľa (2011 PPP \$)	30672,0

Zdroj: vlastné spracovanie podľa [16]

Výsledný index ľudského rozvoja (HDI) pre SR sa vypočíta ako geometrický priemer z hodnôt indexu zdravia, indexu vzdelania a príjmového indexu:

² V roku 2007 dosiahlo Luxembursko veľmi vysokú hodnotu HND (91 519 dolárov na obyvateľa v PPP \$ 2011), t. j. nad hranicou 75 000 dolárov na obyvateľa.

$$HDI = \sqrt[3]{Index\ zdravotia * Index\ vzdelania * Príjmový\ index} = \quad (8)$$

$$= (0,883 * 0,823 * 0,865)^{1/3} = 0,857$$

Hodnoty HDI a jeho jednotlivých komponentov pre Slovensko majú od začiatku merania v roku 1990 až po súčasnosť rastúci trend (tabuľka č. 4). Najväčší nárast môžeme pozorovať pri indexe vzdelania. V roku 2019 sa Slovensko v hodnotení HDI nachádzalo na 36. mieste medzi 189 hodnotenými krajinami sveta na základe údajov z roku 2018 ([18]).

Tabuľka č. 4: Vývoj HDI a jeho komponentov na Slovensku (1990 – 2018)

Rok	Index vzdelania	Index zdravia	Príjmový index	HDI
1990	0,679	0,788	0,753	0,739
2000	0,712	0,820	0,762	0,763
2007	0,779	0,842	0,823	0,814
2010	0,802	0,854	0,832	0,829
2012	0,824	0,863	0,838	0,842
2016	0,822	0,877	0,855	0,851
2017	0,824	0,880	0,859	0,854
2018	0,824	0,883	0,865	0,857

Zdroj: spracované podľa [16]

Všetky krajiny sú každoročne klasifikované v oblasti stupňa ľudského rozvoja a zatriedené do jednej zo štyroch skupín podľa skóre HDI (tabuľka č. 5). Klasifikačná schéma sa vzťahuje aj na každý z troch čiastkových indexov.

Tabuľka č. 5: Klasifikácia ľudského rozvoja podľa UNDP

Hodnota HDI	Skupina rozvoja
0,800 a viac	Veľmi vysoký stupeň ľudského rozvoja
0,700 – 0,799	Vysoký stupeň ľudského rozvoja
0,550 – 0,699	Stredný stupeň ľudského rozvoja
menej ako 0,550	Nízky stupeň ľudského rozvoja

Zdroj: spracované podľa [19]

V článku hodnotíme ľudský rozvoj na základe HDI a jeho komponentov vo vybraných európskych krajinách (spolu 34 krajín). Z popisnej štatistiky indexov (tabuľka č. 6) a z údajov v prílohách článku (príloha č. 2) je zrejmé, že skoro všetky európske krajiny sa v roku 2007 zaradili do skupiny s veľmi vysokým stupňom ľudského rozvoja (tabuľka č. 5, HDI ≥ 0,8). Do nižšej skupiny, skupiny krajín s vysokým stupňom rozvoja, patrilo len 6 štátov, konkrétne Rumunsko, Čierna Hora, Bulharsko, Srbsko, Albánsko a Turecko. Väčšinou išlo o štáty, ktoré ešte nepatrili do EÚ ale boli kandidátskymi krajinami na vstup do Európskej únie. Najvyššiu hodnotu HDI dosiahlo Nórsko (0,938) a najnižšiu Turecko (0,709).

Tabuľka č. 6: Popisné štatistiky indexov HDI pre 34 vybraných európskych krajín (r. 2007 a 2018)

Štatistika	IE_2007	IH_2007	II_2007	IE_2018	IH_2018	II_2018	HDI 2007	HDI 2018	Rank diff
Min	0.557	0.794	0.674	0.712	0.845	0.727	0.709	0.791	-3
Max	0.915	0.948	1.000	0.946	0.979	0.985	0.938	0.954	6
Medián	0.804	0.908	0.864	0.850	0.941	0.874	0.849	0.884	-1
Priemer	0.796	0.885	0.858	0.847	0.924	0.878	0.845	0.882	0

Vysvetlivky: IE – index vzdelania, IH – index zdravia, II – príjmový index, HDI – index ľudského rozvoja, Rank – poradie krajiny v hodnotení podľa HDI, Rank diff – rozdiel v poradí 2007 vs. 2018.

Zdroj: vlastné spracovanie v SAS EG na základe údajov z prílohy č. 1.

V roku 2018 sa hodnoty HDI v sledovaných krajinách zvýšili a do skupiny s veľmi vysokým stupňom ľudského rozvoja (tabuľka č. 5, HDI \geq 0,8) sa dostali skoro všetky analyzované krajiny. Do nižšej skupiny, skupiny krajín s vysokým stupňom rozvoja, patria už len 2 štáty, konkrétne Srbsko (0,799) a Albánsko (0,791). Turecko dosiahlo hodnotu HDI nad hranicou 0,8 (0,803). Najvyššiu hodnotu HDI dosiahlo opäť Nórsko (0,954).

Záverom tejto časti možno konštatovať, že na základe klasifikácie UNDP, rozdeľujeme vybrané európske krajiny len do dvoch skupín. Zaujímavé bude zistiť, aké skupiny, resp. zhluky krajín sa vytvoria použitím viacrozmerných štatistických metód.

2.2 MERANIE ĽUDSKÉHO ROZVOJA VIACROZMERNÝMI ŠTATISTICKÝMI METÓDAMI

Ľudský rozvoj v krajine je zložený jav. Index ľudského rozvoja ako ukazovateľ na meranie tohto zloženého javu môžeme považovať za integrálny ukazovateľ. Hodnota HDI je výsledkom matematických výpočtov na základe hodnôt merateľných čiastkových ukazovateľov (indikátorov). Na základe metodiky OSN (UNDP) sa ako výsledný vzorec na výpočet HDI používa geometrický priemer z 3 indexov, konštruovaných podľa vzorcov uvedených v predchádzajúcej časti.

V tejto časti článku ukážeme, či by na výpočet HDI neboli vhodné viacrozmerné štatistické metódy ako analýza hlavných komponentov, resp. faktorová analýza a na klasifikáciu krajín do skupín zase zhluková analýza.

Viacrozmerné štatistické metódy sa používajú na zjednodušenie štatistických analýz. Často sa v praxi stretávame s tým, že počiatočný počet analyzovaných znakov (premenných) je veľmi vysoký, čo sťažuje analýzu a interpretáciu výsledkov. Pre zjednodušenie je vhodné nahradiť veľký počet prvotných premenných menším počtom podstatných znakov bez toho, že by sme stratili väčšiu časť informácie. Na riešenie tohto problému boli vytvorené dve príbuzné metódy, a to metóda hlavných komponentov (PCA - principal components analysis) a faktorová analýza (FA – factor analysis).

Obe metódy patria do skupiny metód slúžiacich na analýzu skrytých vzťahov medzi premennými. Cieľom je pochopiť a identifikovať, ako sú premenné prepojené, to znamená, ako sú skorelované. Ak sú premenné skorelované, tak je možné rovnaký objem informácie, ktorý vyjadrujú, vyjadriť menším počtom premenných, čiže môžeme tak znížiť dimenziu. Obe tieto metódy preto vychádzajú z analýzy kovariančnej, resp. korelačnej

matice pôvodných východiskových premenných a pokúšajú sa nájsť skryté, nemerateľné (tzv. latentné) premenné. Tieto nové premenné sa nedajú priamo zmerať, ale majú určitú schopnosť vecnej interpretácie.

2.2.1 Analýza hlavných komponentov

Analýza hlavných komponentov je metódou na tvorbu nových premenných, ktoré sú lineárnymi kombináciami pôvodných p premenných. Maximálny počet nových, už nekorelovaných premenných, ktoré môžeme vytvoriť pomocou PCA, sa rovná počtu pôvodných premenných. Nové umelé premenné nazývame hlavné komponenty (PRIN). Hlavné komponenty sú navzájom už nekorelované. Prvý hlavný komponent je najdôležitejší, lebo vysvetľuje najväčšiu časť celkovej variability údajov a ďalšie hlavné komponenty vysvetľujú postupne zvyšnú časť variability, a to tak, že podiel vysvetlenej variability postupne klesá a na posledný hlavný komponent ostáva len nepatrná časť.

Metóda hlavných komponentov sa vo všeobecnosti používa:

- na identifikáciu odľahlých, resp. vplyvných pozorovaní v údajoch (angl. outliers), ktoré niekedy veľmi silno ovplyvňujú výsledky štatistických analýz;
- na zníženie dimenzie viacrozmernej analýzy;
- na odstránenie závislosti medzi premennými a následné použitie nezávislých hlavných komponentov v iných metódach, napr. v zhlukovej analýze alebo pri tvorbe lineárnych regresných modelov na odstránenie multikolinearity.

Predpokladajme, že súbor p pozorovaných premenných $X_1, X_2, X_3, \dots, X_p$ transformujeme do súboru nových premenných $Y_1, Y_2, Y_3, \dots, Y_p$ (hlavných komponentov) tak, že sú lineárnou kombináciou pôvodných premenných a môžeme ich zapísať nasledujúcim spôsobom:

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\vdots \\ Y_p &= a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned} \tag{9}$$

Koeficienty (váhy, saturácie) a_{ij} každej z týchto lineárnych kombinácií sú odhadované tak, aby spĺňali 5 základných vlastností (pozri [14], kapitola 2).

2.2.2 Faktorová analýza

Faktorová analýza vznikla v psychológii (Ch. Spearman, 1904), kde sa dlho výhradne používala a aj dnes sa často používa. Postupne však prenikla aj do iných vedných odborov. Túto metódu však niektorí štatistickí kritizujú pre jej nejednoznačné riešenie, pre subjektivitu v niektorých jej cieľoch a krokoch, hmlistú interpretáciu a približnosť výsledkov.

Faktorová analýza je podobná metóde analýzy hlavných komponentov, pretože takisto je určená na vytváranie nových premenných a na zníženie dimenzie dát s čo najmenšou stratou informácie. Do určitej miery je možné FA považovať za rozšírenie PCA. Na rozdiel

od PCA však vychádza zo snahy vysvetliť závislosti medzi pôvodnými premennými. Medzi nedostatky PCA patrí, že je závislá od merných jednotiek premenných, neposkytuje jednoznačné kritérium na rozhodnutie, či zvolený počet hlavných komponentov vysvetľuje dostatočné percento celkovej variability, a nezaobrá sa chybovým rozptylom premenných. Prístup FA čiastočne odstraňuje tieto nedostatky PCA, má však iné slabé miesta. FA má veľa subjektívnych aspektov a nejednoznačnosť odhadov faktorových parametrov. Prednosťou FA je jej väčšia všeobecnosť a úspornosť, i keď niektoré odhady vyžadujú splnenie aspoň približného viacrozmerného normálneho rozdelenia. Dôležité je aj určenie počtu spoločných faktorov pred vykonaním FA, ktoré musí vychádzať z hypotéz výskumníka v predmetnej aplikačnej oblasti.

Predpokladajme, tak ako pri PCA, že máme súbor p pozorovateľných premenných (náhodných veličín) X_1, X_2, \dots, X_p , ktoré majú viacrozmerné rozdelenie s p -členným vektorom stredných hodnôt μ_x a s kovariančnou maticou Σ_x s hodnotou p . Všeobecný model FA predpokladá, že existuje q v pozadí stojacich *spoločných faktorov* F_1, F_2, \dots, F_q , ktorých je menej ako p (najlepšie výrazne menej). Tieto faktory umožňujú j -tu pozorovateľnú náhodnú premennú $X_j, j = 1, 2, \dots, p$ vyjadriť týmto spôsobom:

$$X_j = \mu_j + a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jq}F_q + \varepsilon_j \quad (10)$$

kde $\varepsilon_j, j = 1, 2, \dots, p$ sú náhodné (chybové) zložky, označované ako *špecifické faktory* a a_{jk} sú *faktorové váhy (náklady, saturácie, záťaž)*, ktoré vyjadrujú vplyv k -tého spoločného faktora na premennú X_j .

Ak použijeme terminológiu z regresnej analýzy, tak faktorové váhy (saturácie) a_{jk} predstavujú regresné koeficienty medzi pozorovateľnými premennými a nepozorovateľnými faktormi. Pri splnení určitých podmienok sú to vlastne kovariancie medzi nimi. Za predpokladu, že pozorované premenné sú merané v rovnakých merných jednotkách, môžeme faktorové saturácie interpretovať ako príspevok k -tého faktora k j -tej vysvetľovanej premennej. V maticovom tvare môžeme model FA zapísať takto:

$$\begin{aligned} \mathbf{X} &= \mu_x + \mathbf{A}\mathbf{F} + \varepsilon \text{ resp.} \\ \mathbf{X} - \mu_x &= \mathbf{A}\mathbf{F} + \varepsilon \end{aligned} \quad (11)$$

kde \mathbf{A} je matica faktorových váh typu $p \times q$, \mathbf{F} je q -členný vektor spoločných faktorov a ε je p -členný vektor špecifických faktorov. Vektor \mathbf{X} je vektorom pôvodných merateľných premenných, ktoré sa nazývajú tiež *indikátory*.

Bez straty všeobecnosti, môžeme vychádzať aj z normovaných premenných Z_j , čo je pri FA bežnejší prípad. Maticový model môžeme potom zapísať v tvare:

$$\begin{aligned} \frac{\mathbf{X} - \mu_x}{\sigma_x} &= \mathbf{A}^*\mathbf{F}^* + \varepsilon^* \\ \mathbf{Z} &= \mathbf{A}^*\mathbf{F}^* + \varepsilon^* \end{aligned} \quad (12)$$

V tomto modeli je rozdiel hlavne v matici faktorových záťaží \mathbf{A}^* , kde prvky α_{jk}^* predstavujú korelačné koeficienty medzi premennými X_j a faktormi F_k . Pre model FA v ktoromkoľvek tvare predpokladáme:

P1. Spoločné faktory F_k , $k = 1, 2, \dots, q$ sú nezávislé a rovnako rozdelené náhodné veličiny s nulovými strednými hodnotami $E(F_k) = 0$ a jednotkovými rozptylmi $D(F_k) = 1$.

P2. Špecifické faktory ε_j , $j = 1, 2, \dots, p$ sú náhodné premenné s nulovými strednými hodnotami $E(\varepsilon_j) = 0$ a s rozptylmi $D(\varepsilon_j) = e_j$, ktoré sú po dvojiciach nezávislé, čiže $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$.

P3. Faktory F_k a ε_j sú nezávislé náhodné premenné pre každú kombináciu $k = 1, 2, \dots, q$ a $j = 1, 2, \dots, p$, čiže $\text{cov}(F_k, \varepsilon_j) = 0$.

Faktorový model $\Sigma_x = \mathbf{A}\mathbf{A}^T + \mathbf{E}$, v ktorom symbol Σ_x označuje kovariančnú maticu, neurčuje maticu faktorových váh \mathbf{A} jednoznačne (okrem prípadu keď $q = 1$). Ak nejaká matica \mathbf{A} vyhovuje FA modelu, tak pre maticu $\mathbf{B} = \mathbf{A}\mathbf{T}$, kde \mathbf{T} je ortogonálna matica (jej stĺpce sú ortogonálne s normou 1), platí: $\Sigma_x = \mathbf{B}\mathbf{B}^T + \mathbf{E}$. Matica \mathbf{B} je tiež riešením FA modelu. Hovoríme, že sme ju získali ortogonálnou rotáciou matice \mathbf{A} .

Matica faktorových váh nám pomáha identifikovať vzťah medzi spoločnými faktormi a identifikátormi. Rotáciou³ faktorov sa snažíme nájsť takú maticu váh, ktorá je prijateľnejšie interpretovateľná. Pretože existuje nekonečne veľa faktorových riešení z jednej kovariančnej, resp. korelačnej matice, vzniká otázka, či existuje *optimálna* množina spoločných faktorov.

Americký vedec Louis L. Thurston (1887 – 1955) vo svojich prácach zaviedol pojem jednoduchej štruktúry, ktorý môže slúžiť ako kritérium na hľadanie optimálneho riešenia, a naformuloval 5 pravidiel pre jednoduchú štruktúru FA modelu:

1. Každý riadok matice faktorových váh by mal obsahovať aspoň jednu nulu.
2. Každý stĺpec matice faktorových váh by mal obsahovať aspoň q núl.
3. Pre každú dvojicu stĺpcov matice faktorových váh by malo byť viacero premenných, ktoré s jedným faktorom majú nulovú váhu, ale s ostatnými faktormi ich majú vysoké.
4. Pre 5 a viac faktorov by už v každej dvojici stĺpcov matice faktorových váh malo byť čo najviac premenných, ktoré majú nulové váhy v oboch stĺpcoch.
5. Pre každú dvojicu stĺpcov matice faktorových váh by malo byť málo premenných, ktoré majú vysoké váhy v oboch stĺpcoch.

Cieľom tohto návodu je (po vylúčení triviálnych faktorov), aby každá korelácia dvoch premenných bola vysvetlená čo najmenším počtom faktorov. Na bežné používanie tohto

³ *Vhodnejší a výstižnejší je pojem transformácia ako rotácia. Ortogonálna transformácia faktorov (tzv. rotácia faktorov) je výpočtová operácia, ktorou sa z matice faktorových váh získava nová matica. Pojem rotácie sa do FA preniesol z geometrického zobrazenia transformácie faktorových váh. Ortogonálna transformácia je geometricky pevná (rigidná) rotácia q súradnicových osí v p -rozmernom priestore.*

návodu je potrebné pojem „nulová“ faktorová váha nahradiť pojmom „malá“ faktorová váha.

Prvé faktorové riešenie získané napr. metódou PCA sa často ani nedá rozumne interpretovať. Je ďalej potrebné tieto faktory transformovať (rotovať) a tak získať zmyslupnejšie (interpretovateľnejšie) faktory. Väčšina metód rotácie sa snaží vo výsledku získať čo najviac faktorových váh blízkyh nule a zároveň čo najviac ostatných (zvyšných) váh blízkyh jednej.

Teória FA a štatistické programy poskytujú celý rad metód transformácie (rotácie) faktorov. Je potrebné sa rozhodnúť, či použijeme ortogonálnu (pravouhlú, kolmú) rotáciu alebo kosouhlú (šikmú) rotáciu. Ortogonálna rotácia vedie k riešeniu s nekorelovanými (nezávislými) faktormi. Prvky matice faktorových váh \mathbf{A} možno interpretovať ako regresné koeficienty závislosti indikátorov od faktorov a tiež ako korelačné koeficienty medzi nimi. Kosouhlá rotácia vedie k získaniu závislých faktorov, čiže navyše poskytuje korelačnú maticu medzi faktormi. Kosouhlú rotáciu niektorí autori odmietajú, kým iní ju vítajú. Tvrdia, že pre prax sú reálnejšie korelované (závislé) spoločné faktory.

Faktorový model vychádzajúci z korelačnej matice indikátorov môžeme zapísať v maticovom tvare nasledovným spôsobom: $\mathbf{R} = \mathbf{A}^* (\mathbf{A}^*)^T + \mathbf{E}^*$.

Predpokladom faktorového modelu je lineárny vzťah medzi indikátormi a faktormi pre jeho jednoduchosť a výhody pri interpretácii faktorových koeficientov ako koeficientov korelácie ako aj predpoklad práve o q počte spoločných faktorov. Pri odhade parametrov FA modelu je veľa štatistických problémov (bližšie pozri [14], kapitola 3). Medzi tie najdôležitejšie patrí nejednoznačnosť FA modelu, odhad parametrov FA modelu, rotácia faktorov a taktiež problém ako určiť čo najmenší počet spoločných faktorov, ktoré by čo najlepšie vysvetľovali koreláciu medzi vstupnými premennými.

Na určenie počtu spoločných faktorov existuje celý rad objektívnych a subjektívnych rád a možností:

1. Zčať FA metódou PCA a určiť počiatočný počet faktorov, ktorý je len orientačný, ale niekde treba začať.
2. Odhadom počtu spoločných faktorov môže byť počet vlastných čísel redukovanej korelačnej matice väčších ako jedna.
3. Niekedy apriórne vieme z iných analýz alebo z teoretickej analýzy problematiky, koľko spoločných faktorov je potrebných na charakterizovanie vzťahov medzi indikátormi.
4. Spoločné faktory by mali vysvetliť čo najviac celkového rozptylu. V exaktných vedách by to malo byť 90 – 95 % a v spoločenských vedách viac ako 60 – 70 %.
5. Spoločné faktory by mali reprezentovať viac ako 90 % celkovej komunality, ktorá je daná súčtom komunalít všetkých p indikátorov.
6. Môžeme použiť, tak ako v PCA, graf „scree plot“, ktorý zobrazuje počet faktorov na osi x a na osi y percento vysvetlenej variability, t. j. hodnoty vlastných čísel redukovanej kovariančnej, resp. korelačnej matice. Za optimálny počet faktorov treba považovať hodnotu na x -ovej osi pred bodom zlomu na krivke vlastných čísel.

7. Do konečného riešenia sa nemajú zahŕňať tzv. triviálne faktory. Triviálne faktory sú také, ktoré významne korelujú len s jedným indikátorom. Lepšie je takýto indikátor z FA vylúčiť a začať znovu. To neznamená, že daná premenná je nepodstatná, ale nehodí sa do faktorovej analýzy a môže sa brať do úvahy samostatne.

Pri odhade parametrov FA modelu rôznymi metódami je podmienka, aby matica **E** (matica špecifických faktorov) bola pozitívne definitná. Niektoré riešenia FA modelu vedú však k nevhodným (nesprávnym) riešeniam. Takéto prípady dostali názov Heywoodove prípady. Heywoodov prípad nastáva vtedy, keď počas iteračného procesu najmenej jeden špecifický faktor je odhadovaný ako nepozitívny (rovný nule alebo záporný). Matica **E** teda nemusí byť nulovou maticou, ale obsahuje najmenej jeden záporný prvok, a to vedie ku komunalitám väčším ako 1, čo nie je vo FA prípustné.

Nie je však pravda, že všetky nevhodné riešenia vyvoláva Heywoodov prípad. Veľa empirických štúdií dokázalo, že príčinami nesprávnych riešení sú:

- malý rozsah vzorky,
- nízky počet vstupných premenných,
- príliš vysoký počet extrahovaných spoločných faktorov,
- príliš nízky počet extrahovaných spoločných faktorov,
- zlá voľba počiatočného odhadu komunalít,
- použitie nevhodného faktorového modelu pre údaje,
- prítomnosť „outlierov“ v dátach.

Faktorovú analýzu môžeme do určitej miery považovať za rozšírenie metódy PCA. Obe metódy majú spoločné, ale aj odlišné vlastnosti, majú svoje výhody, ale aj nedostatky, ktoré môžeme zhrnúť do nasledovných bodov:

- Obe metódy nemá význam použiť, keď pôvodné premenné nie sú korelované, lebo FA nemá čo vysvetliť a PCA vedie k výsledným hlavným komponentom, ktoré sú zhodné s pôvodnými premennými.
- FA sa pokúša vysvetliť kovariancie a korelácie pôvodných premenných pomocou niekoľkých málo spoločných faktorov (tzv. latentných premenných). PCA vysvetľuje len rozptyl premenných.
- Výpočty v PCA sú jednoduchšie a priamočiarejšie. Model FA má veľa predpokladov a výpočty sú náročnejšie a zložitejšie. Bolo vyvinutých veľa metód odhadu parametrov FA modelu a veľa spôsobov rotácie, ktoré vedú k rôznym riešeniam.
- Medzi nedostatky PCA patrí, že je závislá od zmien merných jednotiek premenných. Riešenie FA modelu metódou maximálnej vierohodnosti je invariantné k takýmto zmenám, čo je výhodou FA.

Pri posúdení vhodnosti vstupných údajov pre PCA a FA sa vychádza z korelačnej matice vstupných premenných. Použitie týchto metód vyžaduje významnú vzájomnú koreláciu vstupných premenných. Okrem korelačnej matice je vhodné použiť aj KMO kritérium (Kaiser – Meyer – Olkin), ktoré je založené na porovnaní jednoduchých a parciálnych koeficientov korelácie. KMO sa vypočíta podľa vzorca ([14]):

$$KMO = \frac{\sum_{i \neq j}^p r_{ij}^2}{\sum_{i \neq j}^p r_{ij}^2 + \sum_{i \neq j}^p r_{parc.,ij}^2} \quad (13)$$

KMO štatistika sa počíta ako celková miera adekvátnosti (vhodnosti) výberových dát pre FA (v SAS EG výstupoch ako MSA – Kaiser's measure of sampling adequacy overall) aj ako čiastková miera adekvátnosti jednotlivých indikátorov. Je to miera homogenity premenných. Hodnoty KMO sa netestujú, ale v praxi sa používa tabuľka odporúčaní podľa Kaisera a Ricea (1974) (tabuľka č. 7).

Tabuľka č. 7: FA - Odporúčania pre hodnoty KMO miery

Hodnota KMO štatistiky	Odporúčanie pre adekvátnosť výberových dát
$\geq 0,9$	Vynikajúce
$<0,8; 0,9)$	Chvályhodné
$<0,7; 0,8)$	stredne užitočné
$<0,6; 0,7)$	Priemerné
$<0,5; 0,6)$	Slabé
$<0,5$	Nedostatočné

Zdroj: [14] podľa Kaisera a Ricea (1974)

2.2.3 Zhluková analýza

Zhluková analýza (angl. cluster analysis – CA) zahŕňa širokú škálu metód a postupov, ktoré sa používajú pri riešení problémov typológie objektov a ich klasifikácie. Cieľom zhlukovej analýzy je rozklad súboru objektov opísaných viacerými ukazovateľmi na niekoľko relatívne rovnorodých podmnožín (zhlukov, tried, segmentov) tak, aby objekty patriace do toho istého zhluku si boli čo najviac podobné a objekty patriace do rôznych zhlukov si boli podobné čo najmenej.

Prvotnú úlohu zhlukovej analýzy môžeme matematicky sformulovať takto: ide o zoskupenie objektov X_i ($i = 1, 2, \dots, n$) do zhlukov C_1, C_2, \dots, C_q ($2 \leq q \leq n$) tak, aby objekty patriace do toho istého zhluku si boli blízke, podobné a objekty patriace do rôznych zhlukov si boli vzdialené, odlišné. V porovnaní s inými metódami sú predmetom zhlukovej analýzy objekty (štatistické jednotky, pozorovania), a nie premenné (štatistické znaky, ukazovatele). Neskôr sa metódy zhlukovania použili aj na premenné, čo znamenalo aplikáciu špecifického prístupu faktorovej analýzy v tejto oblasti. Zhlukovanie premenných sa dnes využíva v data miningu, kde sa často stretávame s veľkým počtom skorelovaných vstupných premenných.

Pri aplikovaní zhlukovej analýzy v praxi musí analytik riešiť tieto problémové okruhy:

- výber miery podobnosti alebo vzdialenosti štatistických jednotiek (objektov),
- výber druhu zhlukovacieho postupu,
- výber zhlukovacej metódy,
- určenie počtu významných zhlukov,
- interpretácia zhlukov.

V teórii sú definované 4 skupiny mier podobnosti medzi objektmi (bližšie pozri [14], kapitola 5). V počítačových štatistických programoch je najčastejšie implementovaná euklidovská vzdialenosť.

Podľa spôsobu zoskupovania objektov do zhlukov rozoznávame hierarchické (aglomeračné a divízne) a nehierarchické postupy. V procese hierarchického zhlukovania nie je potrebné dopredu poznať optimálny počet zhlukov. Ten sa určuje dodatočne na základe grafu dendrogramu alebo rôznych mier homogenity, resp. heterogenity zhlukovania. Medzi miery homogenity patria semiparciálny koeficient determinácie (semipartial R-Squared – SPRQ) a vzdialenosť zhlukov (cluster distance – CD). Koeficient determinácie (R-Squared – RSQ) je zase miera heterogenity zhlukov. Čím je táto miera bližšie k číslu 1, tým vyššia je medziskupinová variabilita.

V prípade nehierarchických postupov musí byť vopred známy (daný) počet zhlukov (zhlukovanie s konštantným počtom zhlukov), pričom tento počet sa môže počas zhlukovania meniť (zhlukovanie s optimalizovaným počtom zhlukov, napr. metóda K-means). V princípe pre malé súbory objektov sú vhodné hierarchické postupy a pre veľké objemy zase nehierarchické postupy zhlukovania. Tieto postupy sa navzájom dopĺňujú.

Pre oba zhlukovacie postupy bolo vyvinutých niekoľko zhlukovacích metód, ktoré súvisia so zvoleným spôsobom vyjadrenia vzdialenosti. Pri hierarchických postupoch zhlukovania môžeme použiť napr. tieto metódy: metódu najbližšieho suseda (angl. single linkage), metódu najvzdialenejšieho suseda (angl. complete linkage), metódu priemernej vzdialenosti (angl. average linkage), centroidnú metódu (angl. centroid method), mediánovú metódu (angl. median method) a Wardova metóda (angl. Ward's minimum variance method). Wardova metóda je obľúbená u štatistikov, lebo neoptimalizuje vzdialenosť medzi zhlukmi, ale minimalizuje heterogenitu (rozptyl) vo vnútri zhlukov. Vyžaduje vyjadrenie vzdialenosti objektov pomocou štvorca euklidovskej vzdialenosti. Pri použití rôznych zhlukovacích postupov a metód dostávame rôzne výsledky. Nedá sa všeobecne povedať, ktorá z uvedených metód je najlepšia. Všetky sú však citlivé na prítomnosť extrémnych hodnôt (angl. outliers) v dátach.

Zhluková analýza vyžaduje splnenie určitých predpokladov. Práve na základe vlastností dátovej množiny sa rozhodujeme, ktorá zhlukovacia metóda je pre konkrétny prípad najvhodnejšia. Dáta by nemali obsahovať odľahlé pozorovania (angl. outliers) a chýbajúce hodnoty. Premenné musia byť často štandardizované (normované), aby sa odstránil vplyv rôznych merných jednotiek. Ak by sme premenné nenormovali, prejavili by sa v našej zhlukovej analýze s rôznou dôležitosťou, ktorá by sa odvíjala od ich merných jednotiek, nie charakteru problému. Po odstránení uvedených problémov vyberáme vhodnú metódu na základe rozsahu dátového súboru z hľadiska počtu pozorovaní a podielu počtu premenných k počtu pozorovaní, charakteru premenných (intervalové alebo poradové premenné) a prípadnej apriórnej informácie o počte zhlukov.

Výsledky zhlukovej analýzy negatívne ovplyvňuje aj závislosť medzi vstupnými premennými. Tento jav je možno eliminovať použitím PCA alebo FA, ktoré vytvárajú nové nezávislé premenné (hlavné komponenty alebo spoločné faktory). Takisto je možné

riešenie, že ako vzdialenosť objektov vyberieme Mahalanobisovu vzdialenosť, ktorá dokáže odstrániť vplyv korelácie medzi premennými. V softvéri SAS EG je však možno použiť len euklidovskú vzdialenosť, ktorá to nedokáže.

2.2.4 Použitie analýzy hlavných komponentov a faktorovej analýzy na výpočet HDI

Pri výpočte HDI pomocou PCA je potrebné vychádzať z pôvodných čiastkových ukazovateľov a tie musia byť významne korelované. Ako uvádzame vyššie (tabuľka č. 1), odborníci na skúmanie problematiky v oblasti ľudského rozvoja ako čiastkové indikátory na výpočet HDI vybrali nasledujúce 4 merateľné ukazovatele (indikátory):

- očakávaná dĺžka života pri narodení (LE),
- očakávaný počet rokov štúdia (Exp_Edu),
- priemerný počet rokov štúdia (Mean_Edu) a
- hrubý národný produkt na obyvateľa v PPP \$ v cenách roku 2011 (GNI).

Na účely článku sme použili údaje o týchto štyroch merateľných ukazovateľoch za 34 európskych krajín, ktoré sú dostupné na webovej stránke UNDP [16]. Analyzovali sme dva roky, rok 2007 a 2018. Na štatistické výpočty bol použitý softvér SAS Enterprise Guide.

Popisná štatistika 4 vstupných ukazovateľov (indikátorov) v rokoch 2007 a 2018 za 34 krajín EÚ je v tabuľke č. 8. Z výsledkov je zrejmé, že hodnoty sledovaných ukazovateľov v roku 2018 oproti roku 2007 v analyzovaných 34 európskych krajinách vzrástli. Priemerná očakávaná dĺžka života sa zvýšila zo 77,5 na 80 rokov, t. j. o 2,5 roka. Vzrástli aj oba ukazovatele počtu rokov štúdia. Vzrast HDP na obyvateľa (GNI) za toto obdobie bol v priemere až o 2 894 PPP \$ v cenách roku 2011. Je potrebné si však všimnúť, že maximálna hodnota HDP v roku 2007 bola až 91 519 a dosiahlo ju Luxembursko. Tieto výsledky signalizujú, že kvalita života v krajinách za sledované obdobie vzrástla.

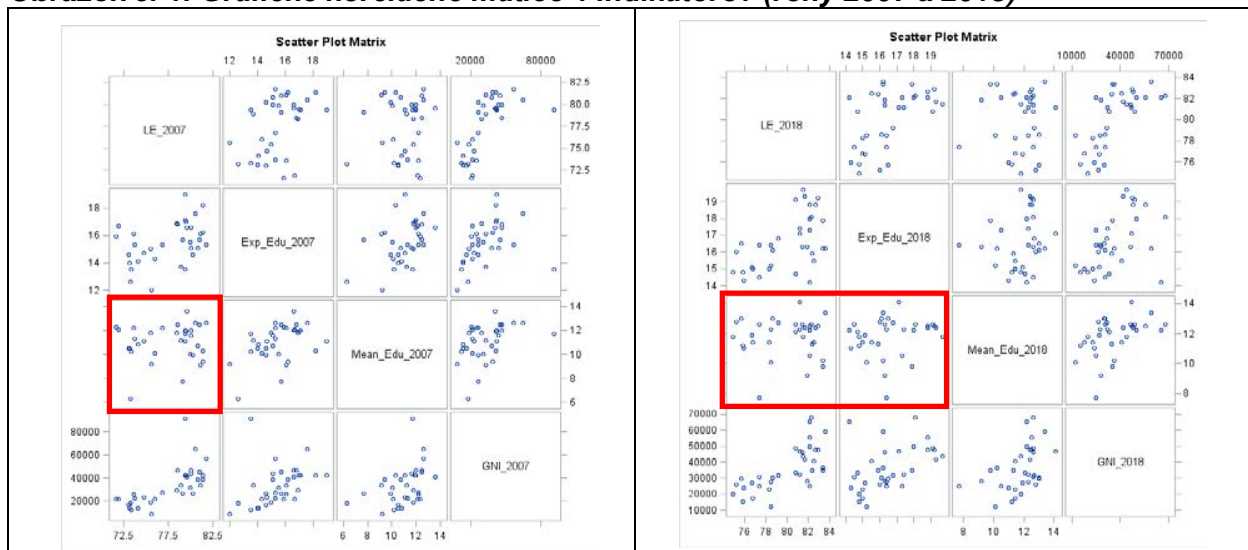
Tabuľka č. 8: Hodnoty popisných štatistík 4 indikátorov pre 34 vybraných európskych krajín (roky 2007 a 2018)

Rok 2007	N	Mean	Std Dev	Median	Min	Max	Rok 2018	N	Mean	Std Dev	Median	Min	Max
LE	34	77.5	3.2	79.0	71.6	81.7	LE	34	80.0	2.7	81.2	74.9	83.6
Exp_Edu	34	15.5	1.5	15.4	12.0	19.0	Exp_Edu	34	16.6	1.6	16.3	14.2	19.7
Mean_Edu	34	11.0	1.5	11.3	6.3	13.6	Mean_Edu	34	11.8	1.3	12.2	7.7	14.1
GNI	34	33177	16850	30363	8671	91519	GNI	34	36071	13953	32622	12300	68059

Zdroj: vlastné spracovanie v SAS EG

Na základe výsledkov korelačných matíc 4 indikátorov (obrázok č. 1) môžeme konštatovať, že v roku 2007 nie je významná korelácia medzi dvoma ukazovateľmi: očakávanou dĺžkou života pri narodení (LE) a priemerným počtom rokov štúdia (Mean_Edu) ($r = 0,11$ a p -hodnota = 0,537). V roku 2018 až 2 korelačné koeficienty nie sú významné, okrem nevýznamnej korelácie medzi očakávanou dĺžkou života pri narodení (LE) a priemerným počtom rokov štúdia (Mean_Edu) ($r = 0,036$ a p -hodnota = 0,842), je aj nevýznamná korelácia medzi priemerným počtom rokov štúdia (Mean_Edu) a očakávaným počtom rokov štúdia (Exp_Edu), ($r = 0,181$ a p -hodnota = 0,307). Toto sa prejaví vo veľkosti vysvetlenej variability hlavnými komponentmi, resp. faktormi.

Obrázok č. 1: Grafické korelačné matice 4 indikátorov (roky 2007 a 2018)



Zdroj: vlastné spracovanie v SAS EG

Aj na základe veľkosti hodnôt celkových KMO mier vidíme, že ukazovatele patria do kategórie slabé na použitie PCA, resp. FA (tabuľka č. 7). Celková KMO miera dosahuje hodnoty len tesne nad 0,5 (tabuľka č. 9: KMO = 0,504 za r. 2007 a KMO = 0,523 za r. 2018). Spôsobili to hlavne nízke hodnoty individuálnych KMO mier pre 2 ukazovatele (sú pod hodnotou 0,5): očakávaná dĺžka života (LE) a priemerný počet rokov štúdia (Mean_Edu).

Tabuľka č. 9: Hodnoty KMO mier pre 4 indikátory (roky 2007 a 2018)

Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.504				Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.523			
LE_2007	Exp_Edu_2007	Mean_Edu_2007	GNI_2007	LE_2018	Exp_Edu_2018	Mean_Edu_2018	GNI_2018
0.469	0.614	0.414	0.530	0.498	0.820	0.301	0.530

Zdroj: vlastné spracovanie v SAS EG

Pri PCA sme vychádzali z korelačnej matice, lebo vstupné indikátory sú vyjadrené v rôznych merných jednotkách. Z dosiahnutých výsledkov v SAS EG vidíme, že 1. hlavný komponent (PRIN1), ktorého vlastné číslo je vyššie ako 1 (t. j. priemer všetkých vlastných čísiel (angl. Eigenvalues)), vysvetľuje len necelých 60 % celkovej variability 4 indikátorov (obrázok č. 2 a tabuľka č. 10: 56,1 % za r. 2007 a 57,3 % za r. 2018). V roku 2018 aj vlastné číslo pre 2. hlavný komponent (PRIN2) je ešte vyššie ako 1 (1,006) a vysvetľuje ešte vysoký podiel variability premenných (až 25,2 %).

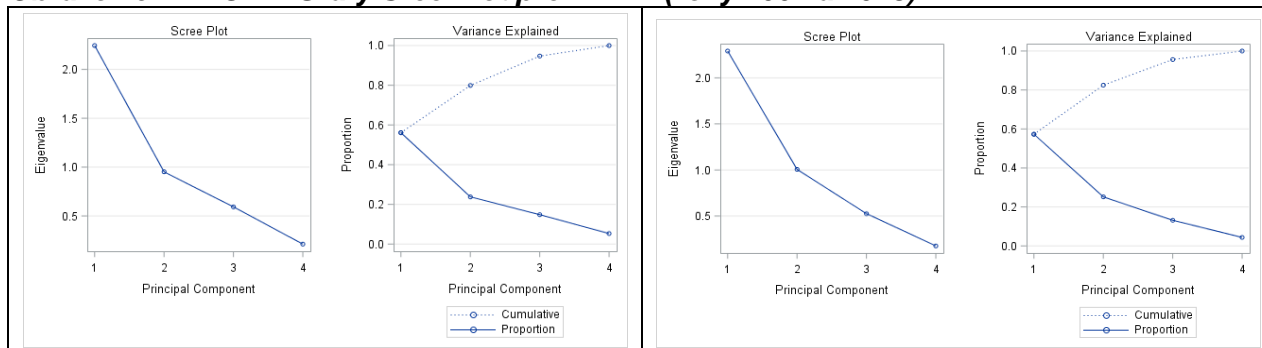
Aj na základe odhadnutých komponentových váh (tabuľka č. 11) je zrejmé, že PCA metóda nie je najlepším riešením na konštrukciu integrálneho ukazovateľa pre ľudský rozvoj, čiže na výpočet HDI. Komponentové váhy PRIN1 pre všetky vstupné ukazovatele sa pohybujú len okolo hodnoty 0,5 a v roku 2018 sa vyskytla veľmi nízka váha pre indikátor Mean_Edu (len 0,3). V ďalších stĺpcoch sa vyskytujú zase vysoké komponentové váhy pre PRIN2 až PRIN4 a aj tieto hlavné komponenty vysvetľujú ešte vysoký podiel variability údajov (nad 5 %, pozri tabuľku č. 10).

Tabuľka č. 10: PCA – hodnoty vlastných čísiel (roky 2007 a 2018)

Eigenvalues of the Correlation Matrix (2007)					Eigenvalues of the Correlation Matrix (2018)				
	Eigenvalue	Difference	Proportion	Cumulative		Eigenvalue	Difference	Proportion	Cumulative
1	2.244	1.293	0.561	0.561	1	2.292	1.286	0.573	0.573
2	0.951	0.359	0.238	0.799	2	1.006	0.480	0.252	0.825
3	0.593	0.380	0.148	0.947	3	0.526	0.351	0.132	0.956
4	0.212		0.053	1.000	4	0.175		0.044	1.000

Zdroj: vlastné spracovanie v SAS EG

Obrázok č. 2: PCA – Grafy Scree Plot pre PRIN1 (roky 2007 a 2018)



Zdroj: vlastné spracovanie v SAS EG

Tabuľka č. 11: PCA – hodnoty komponentových váh (roky 2007 a 2018)

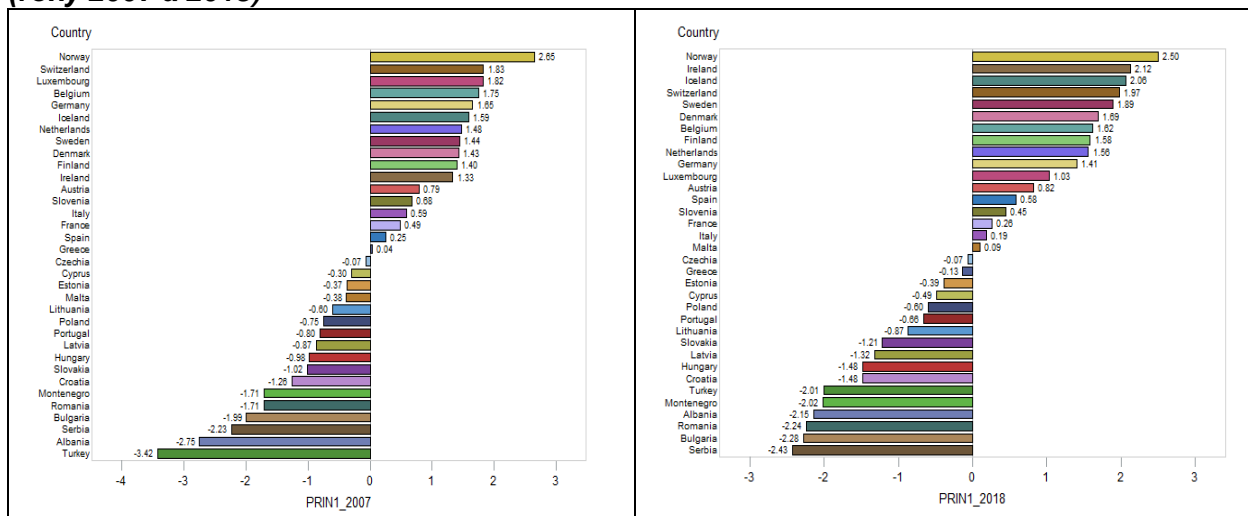
	Eigenvectors					Eigenvectors			
	PRIN1	PRIN2	PRIN3	PRIN4		PRIN1	PRIN2	PRIN3	PRIN4
LE_2007	0.517	-0.572	-0.091	0.630	LE_2018	0.546	-0.411	-0.390	0.617
Exp_Edu_2007	0.491	0.297	-0.781	-0.246	Exp_Edu_2018	0.505	-0.232	0.828	-0.078
Mean_Edu_2007	0.409	0.721	0.412	0.379	Mean_Edu_2018	0.300	0.872	0.096	0.376
GNI_2007	0.570	-0.254	0.461	-0.631	GNI_2018	0.598	0.134	-0.392	-0.687

Zdroj: vlastné spracovanie v SAS EG

Ak aj napriek tomu použijeme hodnoty komponentových skóre PRIN1 pre analyzované krajiny na zostavenie ich poradia podľa úrovne ľudského rozvoja, výsledok môžeme vidieť na obrázku č. 3. V roku 2007 dosiahlo najvyššiu hodnotu Nórsko (kladná hodnota PRIN1 = 2,65) a najnižšiu hodnotu Turecko (záporná hodnota PRIN1 = -3,42). V roku 2018 sa poradie zmenilo hlavne na konci rebríčka. Nórsko síce opäť dosiahlo najvyššiu hodnotu (kladná hodnota PRIN1 = 2,50), ale najnižšiu hodnotu malo Srbsko (záporná hodnota PRIN1 = -2,43). Turecko si polepilo svoju pozíciu v oblasti ľudského rozvoja (záporná hodnota PRIN1 = -2,01). Záverom môžeme konštatovať, že výsledné poradie 34 krajín na základe hodnôt PRIN1 (metóda PCA) ktoré vychádza zo 4 merateľných indikátorov je podobné aké bolo na základe hodnôt HDI poskytnutých metodikou UDNP (pozri príloha č. 2). Len poradie Luxemburska v roku 2007 je výrazne iné (3. miesto), lebo výsledné hodnoty PRIN1 ovplyvnila hlavne extrémne vysoká hodnota ukazovateľa hrubý domáci produkt (GNI).

Pri výpočte HDI pomocou faktorovej analýzy z pôvodných 4 čiastkových ukazovateľov by sme dostali podobné výsledky ako pri PCA. Prvý spoločný faktor však vysvetľuje len necelých 60 % závislosti premenných. Výsledky v článku neuvádzame.

Obrázok č. 3: PCA – poradie 34 krajín podľa hodnôt komponentových skóre pre PRIN1 (roky 2007 a 2018)



Zdroj: vlastné spracovanie v SAS EG

Ak ako vstupné premenné pre PCA, resp. FA použijeme 3 indexy (dimenzie), ktoré boli vypočítané podľa metodiky UDNP, tak postup aj výsledok bude iný. Ako je zrejme z korelačných matíc (tabuľka č. 12, obrázok č. 4), indexy sú skorelované, ale aj medzi nimi sa objavila slabá, resp. nevýznamná korelácia: konkrétne medzi indexom vzdelania (IE) a indexom zdravia (IH) za obidva sledované roky (2007: $r = 0,288$ a 2018: $r=0,372$). Môže to znamenať, že vstupné premenné nie sú vhodné na použitie PCA, resp. FA. Hodnoty celkových aj individuálnych KMO mier sú však vysoké (tabuľka č. 13), preto pomocou FA vypočítame z nich integrálny ukazovateľ HDI ako prvý spoločný faktor (Factor1). Uprednostníme použitie FA namiesto PCA, lebo výsledky FA sa dajú v SAS EG rotovať a to nám umožní lepšiu interpretáciu HDI.

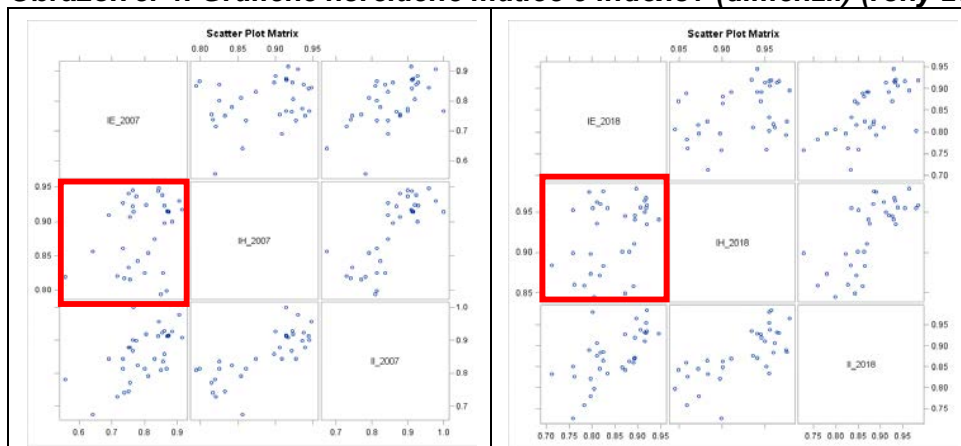
Tabuľka č. 12: Hodnoty Pearsonových koeficientov korelácie medzi 3 indexmi (dimenziami) na výpočet HDI pre 34 vybraných európskych krajín (roky 2007 a 2018)

Pearson Correlation Coefficients, N = 34			
Prob > r under H0: Rho=0			
	IE_2007	IH_2007	II_2007
IE_2007	1	0.2880	0.5949
IH_2007	0.2880	1	0.7803
II_2007	0.5949	0.7803	1
	0.0986	<.0001	<.0001

Pearson Correlation Coefficients, N = 34			
Prob > r under H0: Rho=0			
	IE_2018	IH_2018	II_2018
IE_2018	1	0.3716	0.6576
IH_2018	0.3716	1	0.7187
II_2018	0.6576	0.7187	1
	0.0305	<.0001	<.0001

Zdroj: vlastné spracovanie v SAS EG

Obrázok č. 4: Grafické korelačné matice 3 indexov (dimenzií) (roky 2007 a 2018)



Vysvetlivky: IE – index vzdelania, IH – index zdravia, II – príjmový index

Zdroj: vlastné spracovanie v SAS EG

Tabuľka č. 13: Hodnoty KMO mier pre 3 indexy (roky 2007 a 2018)

Kaiser's Measure of Sampling Adequacy:		
Overall MSA = 0.760		
IE_2007	IH_2007	II_2007
0.8801	0.7465	0.6801

Kaiser's Measure of Sampling Adequacy:		
Overall MSA = 0.779		
IE_2018	IH_2018	II_2018
0.8544	0.7851	0.7103

Zdroj: vlastné spracovanie v SAS EG

Pretože vstupné indexy sú bezrozmerné ukazovatele, tak pri výpočtoch PCA, resp. FA je potrebné vychádzať z kovariančnej matice. Hodnoty vlastných čísel z kovariančnej matice sú uvedené v ďalšej tabuľke. Je zrejmé, že 1. spoločný faktor (Factor1) vysvetľuje takmer 100 % variability údajov (tabuľka č. 14: 2007: 99,79 % a 2018: 99,89 % a obrázok č. 5: Graf Scree Plot). Faktorové váhy všetkých 3 indexov sú veľmi vysoké (blízke číslu 1) v tomto spoločnom faktore (tabuľka č. 15).

Takto získaný Factor1 bude preto vhodný ako integrálny ukazovateľ HDI. Poradie 34 krajín na základe hodnôt faktorových skóre pre Factor1 sme zobrazili na obrázku č. 6) a hodnoty uvádzame aj v prílohe č. 3. Z výsledkov je zrejmé, že poradie krajín na základe hodnôt HDI vypočítaných metodikou UNDP a hodnôt faktorových skóre pre týmto spôsobom získaný Factor1 sú si veľmi podobné (príloha č. 3).

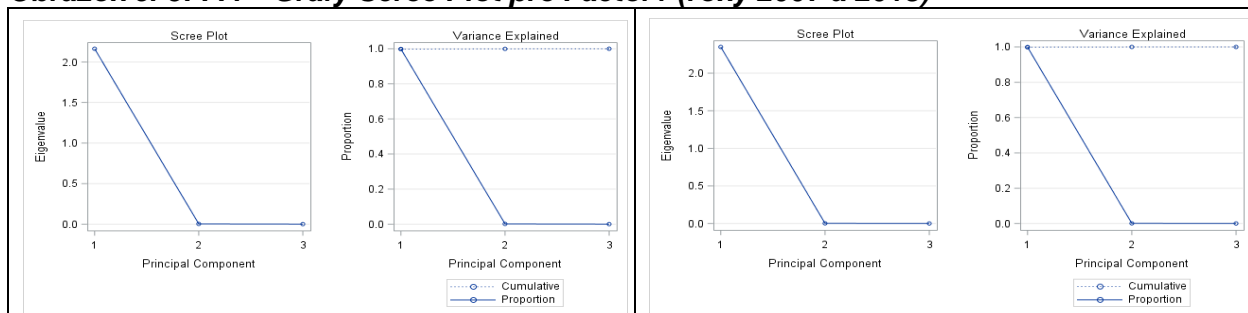
Tabuľka č. 14: Hodnoty vlastných čísel z kovariančnej matice pre 3 indexy (roky 2007 a 2018)

Eigenvalues of the Uncorrected Covariance Matrix (2007)				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.1627	2.1593	0.9979	0.9979
2	0.0034	0.0024	0.0016	0.9995
3	0.0010		0.0005	1.0000

Eigenvalues of the Uncorrected Covariance Matrix (2018)				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.3480	2.3463	0.9989	0.9989
2	0.0018	0.0010	0.0008	0.9996
3	0.0008		0.0004	1.0000

Zdroj: vlastné spracovanie v SAS EG

Obrázok č. 5: FA – Grafy Scree Plot pre Factor1 (roky 2007 a 2018)



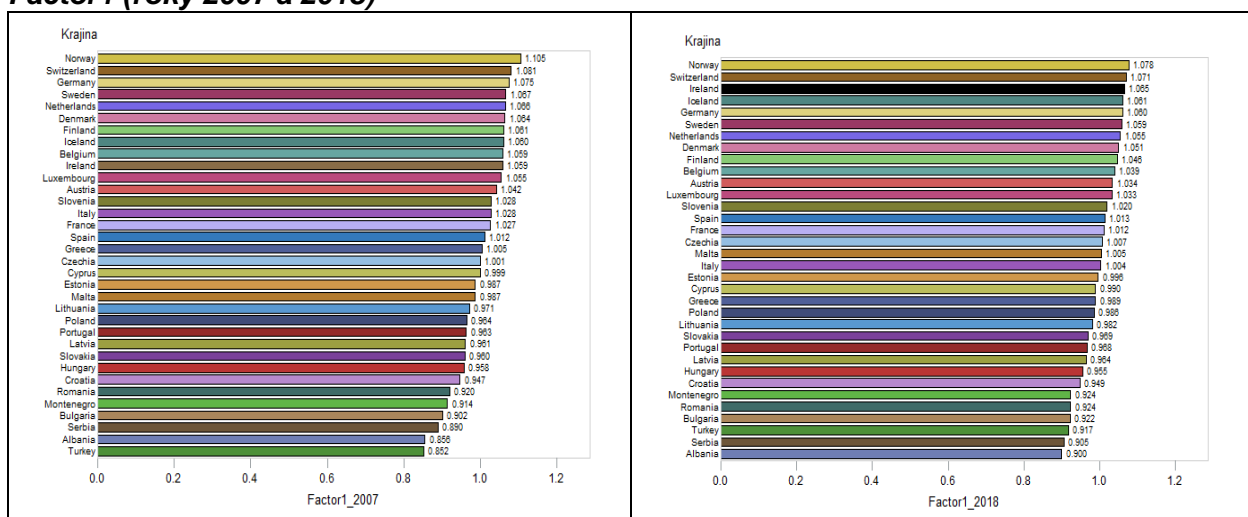
Zdroj: vlastné spracovanie v SAS EG

Tabuľka č. 15: Hodnoty faktorových váh pre Factor1 (roky 2007 a 2018)

Factor Pattern	Factor Pattern	Factor Pattern	Factor Pattern
	Factor1		Factor1
IE_2007	0.9982	IE_2018	0.9992
IH_2007	0.9991	IH_2018	0.9995
II_2007	0.9995	II_2018	0.9996

Zdroj: vlastné spracovanie v SAS EG

Obrázok č. 6: FA – poradie 34 vybraných európskych krajín podľa faktorových skóre pre Factor1 (roky 2007 a 2018)

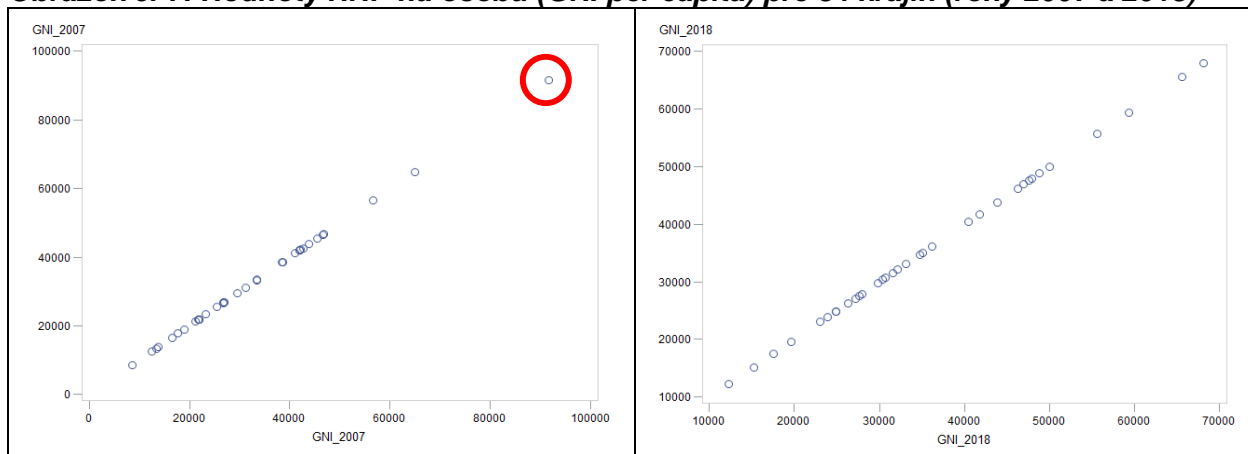


Zdroj: vlastné spracovanie v SAS EG

2.2.5 Použitie zhlukovej analýzy na zaradenie krajín do skupín podľa úrovne HDI

V tejto časti sa pomocou zhlukovej analýzy pokúsime zoskupiť 34 krajín EÚ do vhodného počtu skupín (zhlukov) na základe 4 merateľných ukazovateľov HDI. Zhluková analýza kladie na údaje viacero požiadaviek. Jednou z nich je, aby vstupné premenné neboli navzájom korelované. Druhou požiadavkou je, aby sa v údajoch nevyskytovali extrémne hodnoty (angl. outliers). Veľmi dôležité je aj to, aby vstupné premenné boli vyjadrené v rovnorodých meraciach jednotkách. Na základe predchádzajúcich zistení môžeme konštatovať, že tieto požiadavky naše údaje nespĺňajú, lebo vstupné premenné sú korelované, v dátach sa vyskytujú aj extrémne hodnoty, napríklad hodnota HNP (GNI) v Luxembursku za rok 2007 (obrázok č. 7.) a pôvodné 4 ukazovatele sú rôznorodé.

Obrázok č. 7: Hodnoty HNP na osobu (GNI per capita) pre 34 krajín (roky 2007 a 2018)



Zdroj: vlastné spracovanie v SAS EG na základe údajov:
<http://hdr.undp.org/en/indicators/141706#>

Prvý a tretí problém vieme vyriešiť pomocou PCA alebo FA. Použijeme FA, lebo umožňuje aj rotáciu, čiže nájsť vhodné riešenie z množstva dobrých riešení. Výpočty urobíme z korelačnej matice metódou PCA s ortogonálnou Varimax rotáciou faktorov. Získame tak 4 triviálne nekorelované a normované spoločné faktory v každom sledovanom roku zvlášť (tabuľka č. 16: Factor1 až Factor4).

Tabuľka č. 16: FA – faktorové váhy po ortogonálnej Varimax rotácii (roky 2007 a 2008)

Rotated Factor Pattern					Rotated Factor Pattern				
	Factor1	Factor2	Factor3	Factor4		Factor1	Factor2	Factor3	Factor4
LE_2007	-0.0032	0.9145	0.2086	0.3467	LE_2018	-0.0334	0.9029	0.2843	0.3208
Exp_Edu_2007	0.2249	0.1926	0.9465	0.1285	Exp_Edu_2018	0.0834	0.2502	0.9482	0.1770
Mean_Edu_2007	0.9601	0.0065	0.2152	0.1784	Mean_Edu_2018	0.9797	-0.0135	0.0730	0.1864
GNI_2007	0.2280	0.3879	0.1448	0.8812	GNI_2018	0.3040	0.4182	0.2357	0.8229

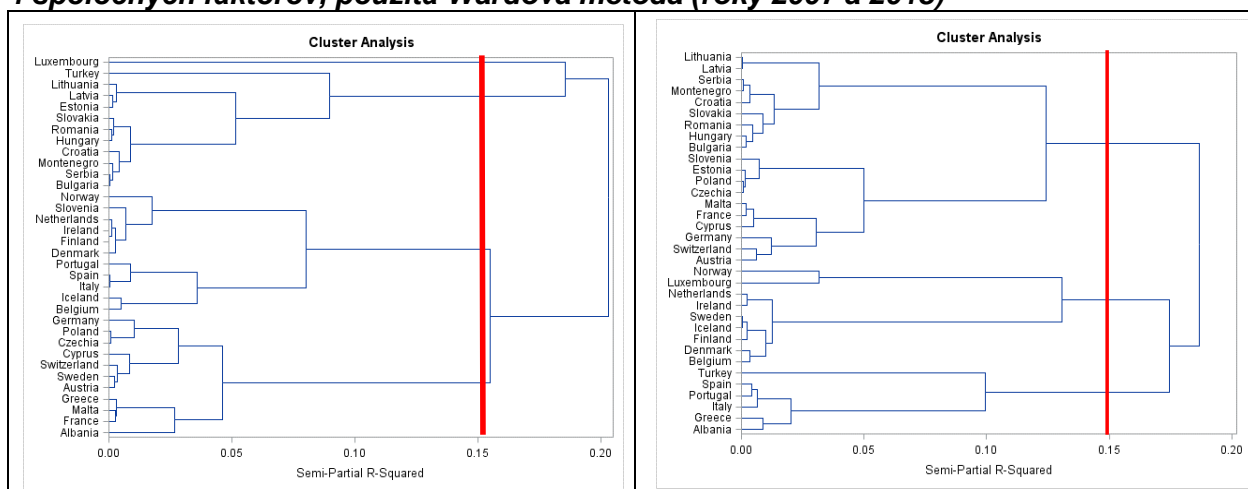
Zdroj: vlastné spracovanie v SAS EG

Tieto nekorelované triviálne faktory použijeme v zhlukovej analýze namiesto 4 merateľných, ale vzájomne skorelovaných ukazovateľov, aby sme získali homogénne skupiny (zhluky) krajín, ktoré sú si navzájom blízke v úrovni ľudského rozvoja.

V SAS EG sme použili z ponúkaných metód hierarchický postup Wardovou metódou. Na výsledných grafoch zhlukovania – dendrogramoch (obrázok č. 8), je zrejmé, že v analyzovaných rokoch 2007 a 2018 sa zhluky krajín líšia. Ak použijeme na určenie počtu zhlukov rovnakú hodnotu štatistiky semi-parciálny koeficient determinácie (SPRQ) na úrovni 0,15, tak v roku 2007 dostaneme 4 zhluky a v roku 2018 len 3 zhluky. V roku 2007 je Luxembursko zaradené do samostatného zhluku, lebo jeho HNP na hlavu (GNI per capita) je vysoké (nad 75 000 \$), čiže ide o extrémnu hodnotu, ktorá negatívne ovplyvnila výsledky zhlukovania. V roku 2018 je už Luxembursko zaradené do zhluku spolu s Nórskom, lebo jeho hodnota HNP sa znížila (pod 75 000 \$).

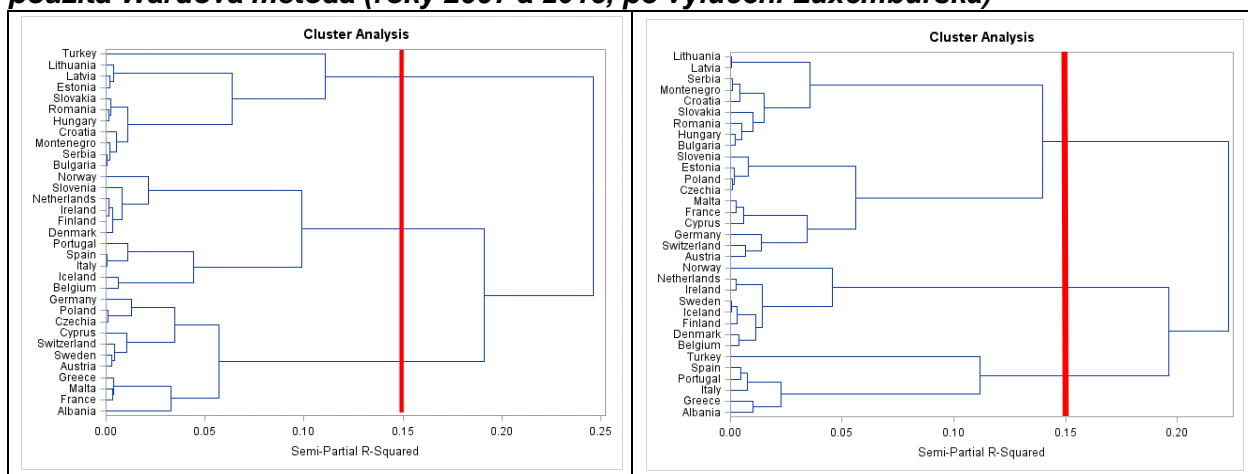
Po vylúčení Luxemburska z analýzy dostávame zmenené dendrogramy (obrázok č. 9). Pri voľbe tej istej hodnoty (0,15) pre štatistiku semi-parciálny koeficient determinácie (SPRQ) dostávame zhodne 3 zhluky za obidva sledované roky.

Obrázok č. 8: CA – dendrogramy zhlukovania 34 vybraných európskych krajín na základe 4 spoločných faktorov, použitá Wardova metóda (roky 2007 a 2018)



Zdroj: vlastné spracovanie v SAS EG

Obrázok č. 9: CA – dendrogramy zhlukovania 33 krajín na základe 4 spoločných faktorov, použitá Wardova metóda (roky 2007 a 2018, po vylúčení Luxemburska)



Zdroj: vlastné spracovanie v SAS EG

S výsledkami zhlukovej analýzy nemôžeme byť spokojní, lebo zhluky sú nelogické. Je to spôsobené výberom len 4 ukazovateľov na meranie úrovne HDI. Druhý dôvod je veľkosť a štruktúra použitých údajov za vybrané krajiny. Ide o pomerne malý súbor štatistických jednotiek ($n = 34$) a premenné majú zošikmené rozdelenia, lebo v každej zo štyroch premenných sa nachádzajú nejaké vybočené, resp. extrémne hodnoty.

Ďalší dôvod je, že sme vylúčili Luxembursko až v zhlukovej analýze. Správny postup je, že extrémne hodnoty treba vylúčiť už pri faktorovej analýze, lebo narušíme normovanie vstupných premenných, čiže faktorov (pozri tabuľka č. 17), ktoré potom používame v zhlukovej analýze. Po úprave výpočtov vo faktorovej analýze (vylúčili sme Luxembursko, $n = 33$) a po opätovnom uskutočnení zhlukovej analýzy s takto upravenými triviálnymi faktormi sme dostali nové dendrogramy (obrázok č. 10).

Tabuľka č. 17: FA – popisná štatistika faktorových skóre pre 34 a 33 krajín (roky 2007 a 2008)

rok 2007	Mean	Std Dev	Min	Max	N	Median
Factor1	0.00	1.00	-3.15	1.79	34	0.22
Factor2	0.00	1.00	-2.11	1.34	34	0.20
Factor3	0.00	1.00	-2.15	2.52	34	-0.10
Factor4	0.00	1.00	-1.30	4.51	34	-0.19

rok 2018	Mean	Std Dev	Min	Max	N	Median
Factor1	0.00	1.00	-3.52	1.84	34	0.13
Factor2	0.00	1.00	-1.87	1.55	34	0.04
Factor3	0.00	1.00	-2.23	2.09	34	-0.05
Factor4	0.00	1.00	-1.58	2.99	34	-0.14

Po vylúčení Luxemburska len v CA

rok 2007	Mean	Std Dev	Min	Max	N	Median
Factor1	0.00	1.02	-3.15	1.79	33	0.23
Factor2	0.02	1.01	-2.11	1.34	33	0.33
Factor3	0.06	0.96	-2.15	2.52	33	0.03
Factor4	-0.14	0.61	-1.30	1.79	33	-0.20

rok 2018	Mean	Std Dev	Min	Max	N	Median
Factor1	0.00	1.02	-3.52	1.84	33	0.13
Factor2	-0.01	1.01	-1.87	1.55	33	0.01
Factor3	0.07	0.93	-1.42	2.09	33	0.06
Factor4	-0.09	0.86	-1.58	2.71	33	-0.20

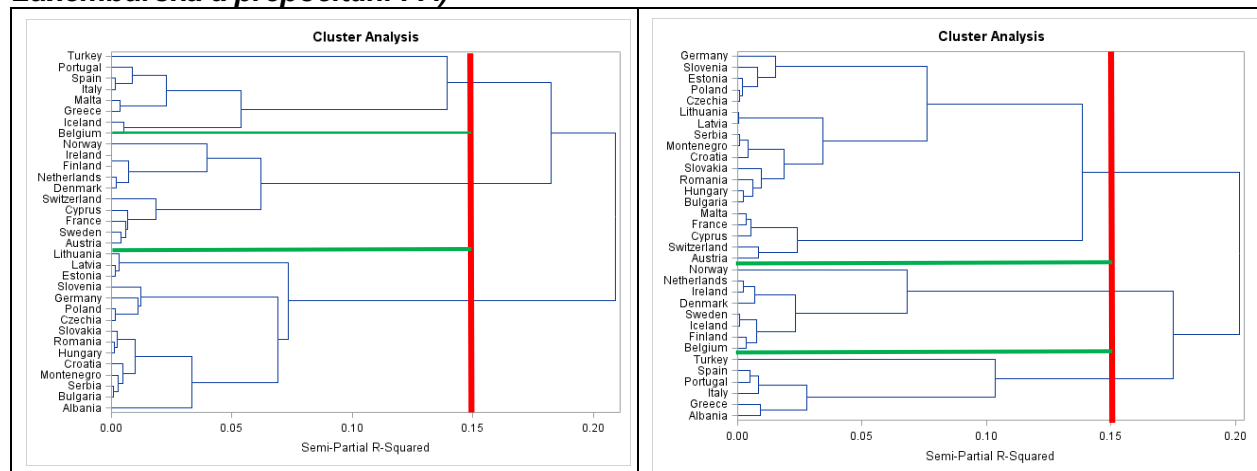
Po vylúčení Luxemburska vo FA, pre novú CA

rok 2007	Mean	Std Dev	Min	Max	N	Median
Factor1	0.00	1.00	-2.01	1.43	33	0.18
Factor2	0.00	1.00	-3.07	1.75	33	0.24
Factor3	0.00	1.00	-2.18	2.60	33	0.01
Factor4	0.00	1.00	-1.46	3.05	33	-0.08

rok 2018	Mean	Std Dev	Min	Max	N	Median
Factor1	0.00	1.00	-1.84	1.54	33	0.02
Factor2	0.00	1.00	-3.47	1.81	33	0.12
Factor3	0.00	1.00	-1.64	2.18	33	-0.06
Factor4	0.00	1.00	-1.67	3.25	33	-0.10

Zdroj: vlastné spracovanie v SAS EG

Obrázok č. 10: CA – dendrogramy zhlukovania 33 európskych krajín na základe 4 spoločných faktorov, použitá Wardova metóda (roky 2007 a 2018, po vylúčení Luxemburska a prepočítaní FA)



Zdroj: vlastné spracovanie v SAS EG

Pri zvolenom počte zhlukov 3 dostávame iné zoskupenia na roky 2007 a 2018 (obrázok č. 10) v porovnaní s výsledkom na obrázku č. 9. Tieto zhluky sú už pomerne rovnorodé a logické z hľadiska úrovne ľudského rozvoja. Zhluky je vhodné interpretovať pomocou popisnej štatistiky pôvodných vstupných ukazovateľov (pozri tabuľka č.18), alebo graficky na základe krabicových grafov (príloha č. 4). Označenie zhlukov číslom 1 až 3 v sledovaných rokoch vytvoril softvér. Zhluky nemožno porovnávať podľa označenia číslom, ale podľa výsledných štatistík. V roku 2007 sú zaradené krajiny s najvyššou úrovňou HDI do zhluku 1 (10 krajín) a v roku 2018 do zhluku 2 (8 krajín). Rozdelenie ukazovateľov v niektorých zhlukoch je pomerne zošíkmené a vyskytujú sa aj extrémne hodnoty, napr. pri ukazovateľoch očakávaná stredná dĺžka života (LE) a hrubý národný produkt na hlavu (GNI) (pozri príloha č. 4).

Tabuľka č. 18: Popisná štatistika výsledných zhlukov pre 33 európskych krajín (roky 2007 a 2018)

CL_2007	N	Variable	Mean	Std Dev	Min	Max	CL_2018	N	Variable	Mean	Std Dev	Min	Max
1	10	LE_2007	80.03	0.99	78.50	81.70	1	19	LE_2018	78.74	2.77	74.9	83.6
		Exp_Edu_2007	16.00	1.21	13.70	17.60			Exp_Edu_2018	15.71	0.93	14.3	17.4
		Mean_Edu_2007	11.96	0.64	10.70	12.60			Mean_Edu_2018	12.23	0.84	11	14.1
		GNI_2007	46066	8939	33342	64859			GNI_2018	31365	10861	15218	59375
2	15	LE_2007	74.73	2.28	71.60	79.60	2	8	LE_2018	82.01	0.67	80.8	82.9
		Exp_Edu_2007	15.00	1.34	12.00	16.90			Exp_Edu_2018	18.88	0.58	18	19.7
		Mean_Edu_2007	11.31	1.12	9.20	13.60			Mean_Edu_2018	12.38	0.27	11.8	12.6
		GNI_2007	21136	8005	8671	41080			GNI_2018	50461	8226	41779	68059
3	8	LE_2007	79.46	2.68	73.20	81.40	3	6	LE_2018	81.12	2.56	77.4	83.4
		Exp_Edu_2007	15.99	1.99	12.60	19.00			Exp_Edu_2018	16.55	0.94	15.2	17.9
		Mean_Edu_2007	9.23	1.54	6.30	11.10			Mean_Edu_2018	9.58	1.02	7.7	10.5
		GNI_2007	32351	8523	17734	42179			GNI_2018	26872	8647	12300	36141

Zdroj: vlastné spracovanie v SAS EG

3. ZÁVER

V príspevku sme opísali princíp výpočtu indexu ľudského rozvoja (HDI) podľa metodiky UNDP. Ľudský rozvoj je zložený jav, ktorý sa priamo nedá zmerať jedným ukazovateľom. Experti vybrali 4 merateľné ukazovatele a vytvorili z nich 3 dimenzie (indexy). Výsledný HDI sa potom vypočíta ako geometrický priemer z týchto 3 indexov.

V článku sme sa pokúsili vypočítať HDI pomocou viacrozmerných štatistických metód, konkrétne pomocou analýzy hlavných komponentov a faktorovej analýzy z pôvodných 4 merateľných ukazovateľov ako aj z jeho 3 dimenzií (indexov). Použili sme údaje z 34 európskych krajín z dvoch rokov, 2007 a 2018. Aj keď použité premenné boli korelované, čo vyžadujú tieto viacrozmerné metódy, údaje obsahovali extrémne hodnoty a to naše riešenie a výsledky negatívne ovplyvnilo. Záverom musíme konštatovať, že výsledky HDI podľa metodiky UNDP sú v princípe lepšie ako sme dostali pomocou viacrozmerných metód, lebo ukazovatele sú škálované (normované) na základe logicky vybraných hodnôt a nielen na základe matematických výpočtov. Podobné výsledky sme dostali len použitím faktorovej analýzy z troch čiastkových indexov (pozri prílohu č. 3). Medzi poradím (Rank) na základe HDI a spoločného faktora (Factor1) sú len minimálne rozdiely v oboch sledovaných rokoch. Toto riešenie pokladáme za vhodnú alternatívu metodiky UNDP na hodnotenie ľudského rozvoja, čiže na výpočet integrálneho ukazovateľa HDI.

Rozvojová komisia OSN (UNDP) každý rok zostavuje poradie 189 krajín sveta v závislosti od veľkosti vypočítaného HDI, čiže sleduje stav a vývoj v oblasti ľudského rozvoja vo svete. UNDP zaraďuje krajiny sveta do 4 skupín podľa úrovne HDI. My sme sa pokúsili pomocou zhlukovej analýzy zoskupiť európske krajiny do zhlukov s podobnou úrovňou ukazovateľov ľudského rozvoja. Aj táto naša snaha bola negatívne ovplyvnená extrémnymi hodnotami použitých ukazovateľov. Výsledné zhluky nie sú rovnorodé a zoskupenia krajín sú nelogické.

Záverom konštatujeme, že metodika UNDP prináša pomerne uspokojivé výsledky pri hodnotení ľudského rozvoja. Výpočet integrálneho ukazovateľa HDI pomocou viacrozmerných štatistických metód narazil na viaceré problémy spojené so štruktúrou údajov a s výskytom extrémnych hodnôt v čiastkových ukazovateľoch.

Cieľom článku bolo metodicky opísať a ukázať, ako pristupovať k hodnoteniu zložitých javov v praxi. Na konštrukciu integrálnych ukazovateľov je v teórii odvodených niekoľko viacrozmerných metód a postupov, ktoré pri dodržaní podmienok ich použitia prinášajú dobré výsledky. V prípade ich nedodržania, výsledky môžu byť skreslené, nelogické až nesprávne.

POĎAKOVANIE

Túto prácu podporili:

- Agentúra na podporu výskumu a vývoja na základe zmluvy č. APVV-16-0091,
- Vedecká grantová agentúra MŠVVaŠ SR a SAV (VEGA) na základe zmluvy č. VEGA 2/0002/19.

LITERATÚRA

- [1] ALTAS, D. – ARIKAN, G. : The Analysis of Human Development Index with Cluster Analysis Techniques. In: Social Sciences Research Journal, 2007, č. 3, s. 126 –138.
- [2] AMALUDDIN, A. et al.: A Modified Human Development Index and Poverty in the Villages of West Seram Regency, Maluku Province, Indonesia. In: International Journal of Economics and Financial Issues, 2018, č. 2, s. 325 – 330.
- [3] BERENGER, V.: Multidimensional measures of well-being: Standard of living and quality of life across countries. In: World Development, 2007, č. 7. s. 1259 – 1276.
- [4] DESAI, M.: Human development: Concept and measurement. In: European Economic Review, 1991, č. 2 – 3, s. 350 – 357.
- [5] DIENER, E. – SUH, E.: Measuring quality of life: economic, social, and subjective indicators. In: Social Indicators Research, 1997, 40, s. 189 – 216.
- [6] KLUGMAN, J. – RODRIGUEZ, F. – CHOI, H. J.: The HDI 2010: new controversies, old critiques. In: The Journal of Economic Inequality, 2011, č. 2, s. 249 – 288.
- [7] KRÁL', P. et al.: Viacrozmerné štatistické metódy so zameraním na riešenie problémov ekonomickej praxe. Banská Bystrica: Univerzita Mateja Bela, Ekonomická fakulta, 2009. 175 s. ISBN 978-80-8083-840-9.
- [8] MAJEROVÁ, I.: Comparison of Old and New Methodology in Human Development and Poverty Indexes: A case of the Least Developed Countries. In: Journal of Economics Studies and Research, 2012, s. 1 – 15.
- [9] MAJEROVÁ, I. – NEVIMA, J.: The measurement of Human Development using the Ward Method of Cluster Analysis. In: Journal of International Studies, 2017, č. 2, 239 – 257.
- [10] MURTAGH, F. – LEGENDRE, P.: Ward's Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm. In: Journal of Classification, 2011, 31, s. 274 – 295.
- [11] OECD: Measuring well-being and progress. Well-being Research. Paris: OECD Statistics and Data, 2019. [online]. [cit. 25.3.2020] Dostupné na: <https://www.oecd.org/statistics/measuring-well-being-and-progress.htm>
- [12] OECD: How's Life? 2020: Measuring Well-being. Paris: OECD Publishing. 2020. [online]. [cit. 25.3.2020] Dostupné na: <https://doi.org/10.1787/9870c393-en>

- [13] SOMARRIBA, N. – PENA, B.: Synthetic indicators of Quality of Life in Europe. In: Social Indicators Research, 2009, č. 1, s.115 – 133.
- [14] STANKOVIČOVÁ, I. – VOJTKOVÁ, M.: Viacrozmerné štatistické metódy s aplikáciami. Bratislava: Iura Edition, 2007. ISBN 978-80-8078-152-1.
- [15] Sustainable Development Goals. [online]. [cit. 25. 3. 2020] Dostupné na: <https://sustainabledevelopment.un.org/sdgs>
- [16] UNDP: Human Development Data (1990 – 2018) [online]. [cit. 25.3.2020] Dostupné na: <http://hdr.undp.org/en/data>
- [17] UNDP Human Development Indices and Indicators: 2018 Statistical Update . 2018. [online]. [cit. 25.3.2020] Dostupné na: http://hdr.undp.org/sites/default/files/2018_human_development_statistical_update.pdf
- [18] UNDP: Human Development Report 2019. Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century. 2019. USA: PBM Graphics, 2014. ISBN 978-92-1-126368-8. [online]. [cit. 25.3.2020] Dostupné na: <http://hdr.undp.org/sites/default/files/hdr2019.pdf>
- [19] UNDP: Technical notes. 2019, [online]. [cit. 25.3.2020] Dostupné na: http://hdr.undp.org/sites/default/files/hdr2019_technical_notes.pdf
- [20] YANG, Y. – HU, A.: Investigating Regional Disparities of China's Human Development with Cluster Analysis: A historical Perspective. In: Social Indicators Research, 2017, č. 3, s. 417 – 432.

RESUMÉ

Článok sa zaoberá metódami výpočtu indexu ľudského rozvoja (HDI) ako integrálneho ukazovateľa zloženého z viacerých čiastkových indikátorov. V prvej časti predstavujeme oficiálnu metodiku výpočtu HDI stanovenú a používanú v Rozvojovom programe Organizácie Spojených národov (UNDP). Takto vypočítaný HDI použijeme na zostavenie poradia vybraných európskych krajín (spolu 34 štátov) za roky 2007 a 2018. V druhej časti článku použijeme na zostavenie poradia týchto krajín v oblasti ľudského rozvoja viacrozmerné štatistické metódy, ako PCA, faktorovú analýzu a zhlukovú analýzu. Výsledky navzájom porovnáme a posúdime výhody a nevýhody oboch prístupov k výpočtu HDI.

RESUME

The paper deals with the methods of calculating the Human Development Index (HDI) as an integral indicator composed of several sub-indicators. In the first part we present the official HDI calculation methodology established and used by the United Nations Development Programme (UNDP). The HDI calculated in this way will be used to compile the ranking of selected European countries (34 countries in total) for years 2007 and 2018. In the second part of the article, we will use multivariate statistical methods such as PCA, factor analysis and cluster analysis to create countries' ranking in human development. We will compare the results and assess the advantages and disadvantages of both approaches for the HDI calculation.

PROFESIJNÝ ŽIVOTOPIS

Mgr. Alena Mojsejová, PhD., ukončila magisterské štúdium na Fakulte prírodných vied UPJŠ v Košiciach. Titul PhD., v odbore ekonometria a operačný výskum, získala na Fakulte hospodárskej informatiky Ekonomickej univerzity v Bratislave. Pôsobí na Ekonomickej fakulte Technickej univerzity v Košiciach. V pedagogickej oblasti sa venuje výučbe pravdepodobnosti a štatistických metód. Vo vedeckej oblasti sa venuje aplikácii štatistických metód v rôznych oblastiach, najmä však v ekonomickej a sociálnej praxi.

Doc. Ing. Iveta Stankovičová, PhD., pôsobí na Fakulte managementu Univerzity Komenského v Bratislave. Dlhodobo sa venuje problematike využitia kvantitatívnych metód v ekonómii a manažmente, čo sa odráža v jej vedeckej profilácii a prednáškovej praxi pre študentov. Vyučuje predmety z oblasti štatistických metód a hĺbkovej analýzy údajov (data mining). Je autorkou vedeckej monografie, spoluautorkou niekoľkých učebníc, skrípt a mnohých vedeckých článkov publikovaných doma i v zahraničí. Od novembra 2014 je predsedníčkou Slovenskej štatistickej a demografickej spoločnosti.

KONTAKT

alena.mojsejova@tuke.sk

iveta.stankovicova@fm.uniba.sk

Príloha č. 1: Hodnoty 4 merateľných ukazovateľov pre 34 vybraných európskych krajín (roky 2007 a 2018)

ID	Krajina	LE 2007	LE 2018	Exp_Edu 2007	Exp_Edu 2018	Mean_Edu 2007	Mean_Edu 2018	GNI 2007	GNI 2018
1	Albania	75.6	78.5	12.0	15.2	9.2	10.1	8671	12300
2	Austria	80.0	81.4	15.1	16.3	11.6	12.6	43896	46231
3	Belgium	79.4	81.5	19.0	19.7	11.1	11.8	42179	43821
4	Bulgaria	73.1	74.9	14.0	14.8	10.5	11.8	13447	19646
5	Croatia	76.0	78.3	14.3	15.0	10.1	11.4	21755	23061
6	Cyprus	78.9	80.8	13.7	14.7	11.2	12.1	33342	33100
7	Czechia	76.8	79.2	15.3	16.8	12.2	12.7	26958	31597
8	Denmark	78.5	80.8	16.8	19.1	12.5	12.6	46609	48836
9	Estonia	73.6	78.6	16.1	16.1	12.2	13.0	25555	30379
10	Finland	79.4	81.7	17.1	19.3	12.0	12.4	42604	41779
11	France	80.8	82.5	15.1	15.5	10.7	11.4	38425	40511
12	Germany	79.6	81.2	16.6	17.1	13.6	14.1	41080	46946
13	Greece	79.9	82.1	15.5	17.3	10.0	10.5	31243	24909
14	Hungary	73.7	76.7	15.3	15.1	11.3	11.9	21944	27144
15	Iceland	81.4	82.9	18.2	19.2	10.3	12.5	42014	47566
16	Ireland	79.5	82.1	17.0	18.8	11.9	12.5	42012	55660
17	Italy	81.4	83.4	16.2	16.2	9.4	10.2	38642	36141
18	Latvia	71.6	75.2	15.9	16.0	12.3	12.8	21284	26301
19	Lithuania	71.9	75.7	16.7	16.5	12.1	13.0	21881	29775
20	Luxembourg	79.4	82.1	13.5	14.2	11.7	12.2	91519	65543
21	Malta	80.2	82.4	14.6	15.9	9.9	11.3	26809	34795
22	Montenegro	74.1	76.8	14.1	15.0	10.8	11.4	13767	17511
23	Netherlands	80.0	82.1	16.6	18.0	12.0	12.2	46861	50013
24	Norway	80.5	82.3	17.6	18.1	12.6	12.6	64859	68059
25	Poland	75.5	78.5	15.0	16.4	11.8	12.3	18903	27626
26	Portugal	79.1	81.9	15.7	16.3	7.7	9.2	26694	27935
27	Romania	73.0	75.9	14.6	14.3	10.5	11.0	16494	23906
28	Serbia	73.3	75.8	13.5	14.8	10.2	11.2	12506	15218
29	Slovakia	74.7	77.4	14.7	14.5	11.1	12.6	23310	30672
30	Slovenia	78.4	81.2	16.9	17.4	11.8	12.3	29483	32143
31	Spain	81.1	83.4	16.1	17.9	9.1	9.8	33494	35041
32	Sweden	81.0	82.7	15.7	18.8	12.5	12.4	45438	47955
33	Switzerland	81.7	83.6	15.3	16.2	12.6	13.4	56611	59375
34	Turkey	73.2	77.4	12.6	16.4	6.3	7.7	17734	24905

Vysvetlivky: LE – očakávaná dĺžka života pri narodení v rokoch, Exp_Edu – očakávaný počet rokov štúdia, Mean_Edu – priemerný počet rokov štúdia, GNI – hrubý národný produkt na obyvateľa v PPP \$ (v cenách roku 2011)

Zdroje údajov: <http://hdr.undp.org/en/content/human-development-index-hdip> a <http://hdr.undp.org/en/indicators/141706#>

Príloha č. 2: Hodnoty čiastkových indexov na výpočet HDI a poradie krajín podľa HDI (roky 2007 a 2018)

ID	Krajina	IE 2007	IH 2007	II 2007	IE 2018	IH 2018	II 2018	HDI 2007	HDI 2018	Rank_HDI 2007	Rank_HDI 2018	Rank Diff
1	Albania	0.642	0.856	0.674	0.758	0.899	0.727	0.718	0.791	33	34	-1
2	Austria	0.806	0.924	0.919	0.871	0.945	0.927	0.881	0.914	12	11	1
3	Belgium	0.871	0.914	0.913	0.893	0.946	0.919	0.899	0.919	8	10	-2
4	Bulgaria	0.738	0.817	0.740	0.805	0.845	0.798	0.764	0.816	31	29	2
5	Croatia	0.734	0.861	0.813	0.796	0.898	0.822	0.801	0.838	28	28	0
6	Cyprus	0.755	0.906	0.878	0.811	0.936	0.876	0.844	0.873	19	20	-1
7	Czechia	0.830	0.874	0.845	0.892	0.911	0.869	0.849	0.891	17	16	1
8	Denmark	0.884	0.900	0.928	0.920	0.935	0.935	0.904	0.930	6	8	-2
9	Estonia	0.856	0.825	0.837	0.881	0.901	0.863	0.839	0.882	20	19	1
10	Finland	0.874	0.914	0.915	0.915	0.950	0.912	0.901	0.926	7	9	-2
11	France	0.775	0.936	0.899	0.811	0.962	0.907	0.867	0.891	14	15	-1
12	Germany	0.915	0.917	0.909	0.946	0.941	0.929	0.914	0.939	3	4	-1
13	Greece	0.764	0.922	0.868	0.833	0.955	0.834	0.849	0.872	18	21	-3
14	Hungary	0.801	0.825	0.814	0.816	0.872	0.846	0.813	0.844	26	27	-1
15	Iceland	0.842	0.944	0.912	0.918	0.967	0.931	0.898	0.938	10	5	5
16	Ireland	0.870	0.915	0.912	0.918	0.955	0.955	0.899	0.943	9	3	6
17	Italy	0.765	0.945	0.900	0.793	0.975	0.890	0.867	0.883	15	18	-3
18	Latvia	0.850	0.794	0.810	0.871	0.849	0.842	0.818	0.854	24	25	-1
19	Lithuania	0.867	0.799	0.814	0.890	0.858	0.860	0.826	0.869	22	23	-1
20	Luxembourg	0.766	0.914	1.000	0.802	0.955	0.980	0.888	0.909	11	12	-1
21	Malta	0.734	0.927	0.845	0.818	0.960	0.884	0.832	0.885	21	17	4
22	Montenegro	0.751	0.833	0.744	0.797	0.873	0.780	0.775	0.816	30	30	0
23	Netherlands	0.862	0.923	0.929	0.906	0.956	0.939	0.904	0.933	5	7	-2
24	Norway	0.907	0.930	0.978	0.919	0.958	0.985	0.938	0.954	1	1	0
25	Poland	0.811	0.854	0.792	0.866	0.901	0.849	0.819	0.872	23	22	1
26	Portugal	0.691	0.909	0.844	0.759	0.952	0.851	0.809	0.850	27	26	1
27	Romania	0.756	0.815	0.771	0.762	0.860	0.827	0.780	0.815	29	31	-2
28	Serbia	0.716	0.820	0.729	0.783	0.859	0.759	0.754	0.799	32	33	-1
29	Slovakia	0.779	0.842	0.823	0.824	0.883	0.865	0.814	0.857	25	24	1
30	Slovenia	0.862	0.898	0.859	0.893	0.941	0.872	0.873	0.902	13	13	0
31	Spain	0.751	0.940	0.878	0.824	0.976	0.885	0.853	0.893	16	14	2
32	Sweden	0.854	0.938	0.924	0.914	0.964	0.932	0.905	0.936	4	6	-2
33	Switzerland	0.844	0.948	0.958	0.896	0.979	0.965	0.915	0.946	2	2	0
34	Turkey	0.557	0.819	0.782	0.712	0.884	0.833	0.709	0.806	34	32	2

Vysvetlivky: IE – index vzdelania, IH – index zdravia, II – príjmový index, HDI – index ľudského rozvoja, Rank – poradie, Rank diff = Rank_2007 – Rank_2018 – zmena v poradí (kladné hodnoty znamenajú zlepšenie postavenia krajiny v rebríčku podľa HDI, resp. záporné hodnoty znamenajú zhoršenie postavenia krajiny).

Zdroj údajov: <http://hdr.undp.org/en/data> a vlastné spracovanie v SAS EG

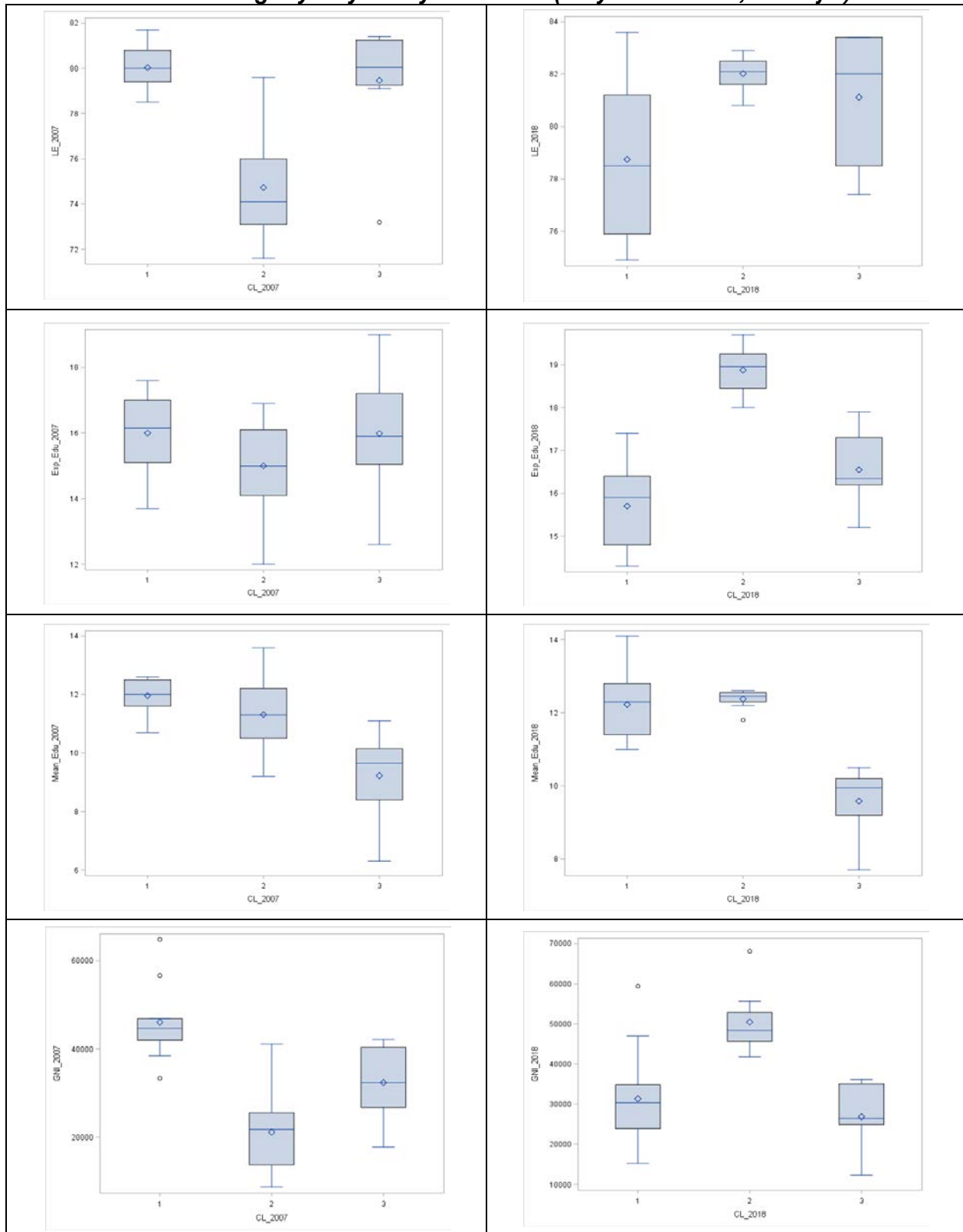
Príloha č. 3: Hodnoty a poradia pre ukazovatele HDI, 1. hlavný komponent (PRIN1) a 1. spoločný faktor (Factor1) pre 34 vybraných európskych krajín (roky 2007 a 2018)

ID	Country	HDI 2007	Rank	HDI 2018	Rank	PRIN1 2007	Rank	PRIN1 2018	Rank	Factor1 2007	Rank	Factor1 2018	Rank
1	Albania	0.718	33	0.791	34	-2.746	33	-2.146	31	0.856	33	0.900	34
2	Austria	0.881	12	0.914	11	0.790	12	0.816	12	1.042	12	1.034	11
3	Belgium	0.899	8	0.919	10	1.745	4	1.618	7	1.059	9	1.039	10
4	Bulgaria	0.764	31	0.816	29	-1.991	31	-2.277	33	0.902	31	0.922	31
5	Croatia	0.801	28	0.838	28	-1.261	28	-1.484	28	0.947	28	0.949	28
6	Cyprus	0.844	19	0.873	20	-0.298	19	-0.487	21	0.999	19	0.990	20
7	Czechia	0.849	17	0.891	16	-0.066	18	-0.068	18	1.001	18	1.007	16
8	Denmark	0.904	6	0.930	8	1.430	9	1.691	6	1.064	6	1.051	8
9	Estonia	0.839	20	0.882	19	-0.369	20	-0.390	20	0.987	20	0.996	19
10	Finland	0.901	7	0.926	9	1.398	10	1.584	8	1.061	7	1.046	9
11	France	0.867	14	0.891	15	0.488	15	0.257	15	1.027	15	1.012	15
12	Germany	0.914	3	0.939	4	1.654	5	1.410	10	1.075	3	1.060	5
13	Greece	0.849	18	0.872	21	0.039	17	-0.134	19	1.005	17	0.989	21
14	Hungary	0.813	26	0.844	27	-0.976	26	-1.480	27	0.958	27	0.955	27
15	Iceland	0.898	10	0.938	5	1.586	6	2.063	3	1.060	8	1.061	4
16	Ireland	0.899	9	0.943	3	1.335	11	2.124	2	1.059	10	1.065	3
17	Italy	0.867	15	0.883	18	0.589	14	0.189	16	1.028	14	1.004	18
18	Latvia	0.818	24	0.854	25	-0.870	25	-1.320	26	0.961	25	0.964	26
19	Lithuania	0.826	22	0.869	23	-0.601	22	-0.867	24	0.971	22	0.982	23
20	Luxembourg	0.888	11	0.909	12	1.821	3	1.027	11	1.055	11	1.033	12
21	Malta	0.832	21	0.885	17	-0.378	21	0.095	17	0.987	21	1.005	17
22	Montenegro	0.775	30	0.816	30	-1.707	29	-2.021	30	0.914	30	0.924	29
23	Netherlands	0.904	5	0.933	7	1.478	7	1.560	9	1.066	5	1.055	7
24	Norway	0.938	1	0.954	1	2.649	1	2.497	1	1.105	1	1.078	1
25	Poland	0.819	23	0.872	22	-0.751	23	-0.597	22	0.964	23	0.986	22
26	Portugal	0.809	27	0.850	26	-0.805	24	-0.663	23	0.963	24	0.968	25
27	Romania	0.780	29	0.815	31	-1.713	30	-2.240	32	0.920	29	0.924	30
28	Serbia	0.754	32	0.799	33	-2.233	32	-2.428	34	0.890	32	0.905	33
29	Slovakia	0.814	25	0.857	24	-1.016	27	-1.215	25	0.960	26	0.969	24
30	Slovenia	0.873	13	0.902	13	0.677	13	0.449	14	1.028	13	1.020	13
31	Spain	0.853	16	0.893	14	0.253	16	0.584	13	1.012	16	1.013	14
32	Sweden	0.905	4	0.936	6	1.439	8	1.890	5	1.067	4	1.059	6
33	Switzerland	0.915	2	0.946	2	1.828	2	1.972	4	1.081	2	1.071	2
34	Turkey	0.709	34	0.806	32	-3.420	34	-2.008	29	0.852	34	0.917	32

Vysvetlivky: HDI – hodnoty indexu ľudského rozvoja (podľa metodiky UNDP), PRIN1 – komponentné skóre pre 1. hlavný komponent (vypočítaný z korelačnej matice 4 vstupných indikátorov), Factor1 – faktorové skóre pre 1. spoločný faktor (vypočítaný z kovariančnej matice 3 dimenzií – indexov), Rank – poradie krajiny v rebríčku podľa hodnoty ukazovateľa v sledovanom roku.

Zdroj údajov: <http://hdr.undp.org/en/data> a vlastné spracovanie v SAS EG

Príloha č. 4: Krabicové grafy 3 výsledných zhlukov (roky 2007 a 2018, 33 krajín)



Zdroj: vlastné spracovanie v SAS EG