

Univerzita Pardubice

Fakulta ekonomicko-správní

Modelování kvality života pomocí faktorové analýzy

Bc. Marie Augustinová

Diplomová práce

2010

Univerzita Pardubice
Fakulta ekonomicko-správní
Akademický rok: 2009/2010

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: Marie AUGUSTINOVÁ
Studijní program: N6209 Systémové inženýrství a informatika
Studijní obor: Regionální a informační management
Název tématu: Modelování kvality života pomocí faktorové analýzy
Zadávací katedra: Ústav systémového inženýrství a informatiky

Z á s a d y p r o v y p r a c o v á n í :

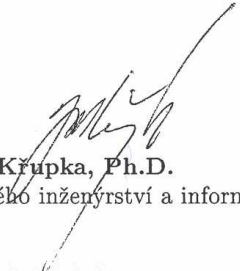
Předpokládá se, že výstupem diplomové práce bude:

- a) úvod do problematiky kvality života pro potřeby definice modelu;
- b) návrh modelu pomocí faktorové analýzy;
- c) analýza modelu.

Rozsah grafických prací:
Rozsah pracovní zprávy:
Forma zpracování diplomové práce: tištěná/elektronická


Seznam odborné literatury:

- MEDERLY, P., TOPERCER, J., NOVÁČEK, P. Indikátory kvality života a udržitelného rozvoje : kvantitativní, vícerozměrný a variantní přístup. Praha: UK FSV CESES. 2004. 117 s. ISBN 80-239-4389-8.
LAROSE, D. T. Data Mining: methods and models. New Jersey: John Wiley & Sons. 2006. ISBN 0-471-66656-4.
O'SULLIVAN, E., RASSEL, G. R., BERNER, M. Research Methods for Public Administrators, 5th Edition. NY: Longman. 2008. ISBN 978-0-321-43137-0.

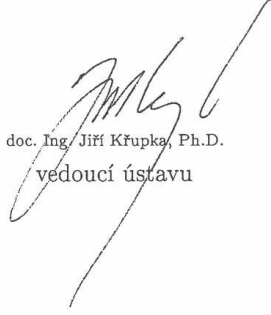
Vedoucí diplomové práce: 
doc. Ing. Jiří Krupka, Ph.D.
Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce: 5. října 2009

Termín odevzdání diplomové práce: 30. dubna 2010


doc. Ing. Renáta Myšková, Ph.D.
děkanka

L.S.


doc. Ing. Jiří Krupka, Ph.D.
vedoucí ústavu

V Pardubicích dne 5. října 2009

Prohlašuji:

Tuto práci jsem vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 26. 4. 2010

Marie Augustinová

Poděkování

Ráda bych touto cestou poděkovala vedoucímu diplomové práce doc. Ing. Jiřímu Křupkovi, Ph.D. za jeho odborné vedení, cenné připomínky a rady. Můj dík patří i rodině za podporu, kterou mi věnovala při studiu.

Anotace

Diplomová práce se zaměřuje na kvalitu života, k jejímu modelování využívá statistickou metodu - faktorovou analýzu. Faktorová analýza odhaluje skryté závislosti v datech. Kvalita života je v práci zkoumána na úrovni okresů České republiky. Byly shromážděny proměnné související s kvalitou života a vytvořeny matice dat. Z těchto dat byly vytvořeny modely, které byly analyzovány a popsány.

Klíčová slova

kvalita života, faktorová analýza, modelování, pozorované proměnné

Title

Modeling quality of life by means of factor analysis

Annotation

This thesis focuses on quality of life for its modeling using a statistical method - factor analysis. Factor analysis reveals a hidden dependence in the data. Quality of life is examined in the work at the district level of the Czech Republic. Were collected variables related to quality of life and created a matrix of data. These data were created models that were analyzed and described.

Keywords

quality of life, factor analysis, modeling, observed variables

Obsah

1	Úvod	8
2	Kvalita života.....	9
2.1	Indikátory kvality života.....	11
2.2	Index lidského rozvoje.....	12
3	Faktorová analýza.....	13
3.1	Základní pojmy.....	14
3.2	Předpoklady faktorové analýzy	17
3.3	Rotace	18
3.4	Metody získávání faktorů	18
4	Formulování problému	20
4.1	Datová matice	21
4.2	Datový slovník.....	22
4.3	Zpracování	28
4.3.1	Metoda hlavních os, matice A	29
4.3.2	Metoda hlavních komponent, matice A.....	32
4.3.3	Metoda hlavních os, matice B	35
4.3.4	Metoda hlavních komponent, matice B.....	37
4.4	Interpretace výstupů	40
4.4.1	Experiment 1	40
4.4.2	Experiment 2	41
4.4.3	Experiment 3	44
5	Závěr.....	46
6	Použité zdroje	49

Seznam obrázků

Obrázek 1 – Faktory které vytvářejí kvalitu života	10
Obrázek 2 – Model kvality života	10
Obrázek 3 – Model faktorové analýzy	16
Obrázek 4 – Navržený model řešení.....	20
Obrázek 5 – Matice faktorových zátěží, experiment A1	30
Obrázek 6 – Matice faktorových zátěží, experiment B1	33
Obrázek 7 – Matice faktorových zátěží, experiment C1	35
Obrázek 8 – Matice faktorových zátěží, experiment D1	38

Seznam tabulek

Tabulka 1 – Datový slovník	23
Tabulka 2 – Odfiltrované proměnné.....	44

1 Úvod

Diplomová práce se zaměřuje na problematiku modelování kvality života. Práce je založena na předpokladu, že pro modelování v dané oblasti lze použít faktorovou analýzu. Cílem práce je, pro tvorbu modelu kvality života využít faktorovou analýzu a analyzovat a interpretovat výstupy modelu.

Začátek práce je věnován přiblížení problematiky kvality života. Jedná se zejména o vymezení pojmu a seznámení s uplatňovanými přístupy. Jsou představeny faktory ovlivňující kvalitu života a problémy a hrozby, které ji mohou výrazně snížit. Současně je zde přiblížen index lidského rozvoje, ukazatel, který používá Organizace spojených národů (OSN) k hodnocení kvality života členských států.

Další část práce je věnována seznámení s faktorovou analýzou. V tomto úseku je ozřejmáno, k čemu se faktorová analýza využívá, jaké dva přístupy se v souvislosti s ní využívají, dále jsou objasněny základní pojmy, s kterými se lze ve faktorové analýze setkat. Nejsou opomenuty ani předpoklady použití faktorové analýzy, možnosti rotace faktorů či metody extrakce faktorů.

Po seznámení s teoretickými východisky je zbývající část práce věnována formulování problému a jeho postupnému zpracování. Součástí zpracování je sestavení matice vstupních hodnot, vytvoření datového slovníku a samotné modelování a v neposlední řadě analýza a interpretace výstupů. Pro zvýšení přehlednosti nebo z důvodu vyššího obsahu informací jsou některá data zobrazena formou tabulek či obrázků.

2 Kvalita života

Kvalita života je složitý a velmi široký pojem. Je těžko uchopitelný pro svou multidimenzionalitu a komplexnost. Výraz kvalita vyjadřuje charakteristický rys toho, o čem se hovoří, je to znak, kterým se předmět odlišuje od jiného ve smyslu lepší nebo horší. Životem je zde rozuměn život lidí [7].

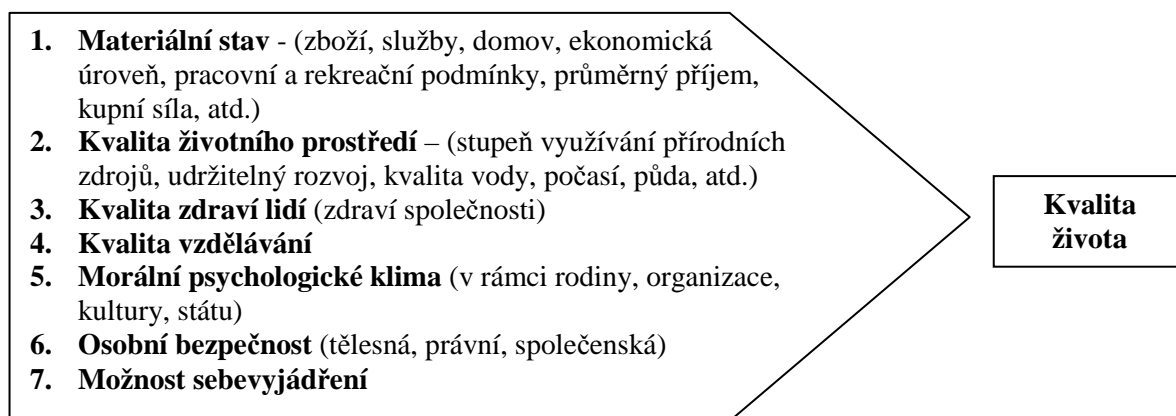
O kvalitě života je možné hovořit ve statickém či dynamickém pojetí. Statické pojetí podává zprávu o životě lidí k určitému časovému údaji, zatímco dynamické pojetí porovnává a hodnotí kvalitu života za delší časové období, např. vývoj kvality života lidí v domech s pečovatelskou službou v posledních patnácti letech.

Významným činitelem při zkoumání kvality života je, kdo se ptá a v jakém kontextu se tento zájemce ptá. Ptát se mohou odborníci různých profesí nebo různé skupiny lidí a podle jejich zájmu bude určováno i to, čemu bude věnována větší pozornost. Je rozdílem, zda odpověď na otázku hledá politik, duchovní, sociolog nebo podnikatel aj., protože zájmem tazatele jsou zcela odlišné odpovědi. Cílem jednoho může být stanovení politického opatření, zatímco druhý se může zabývat výzkumem nebo zavedením ISO standardů pro hodnocení kvality podniku [7].

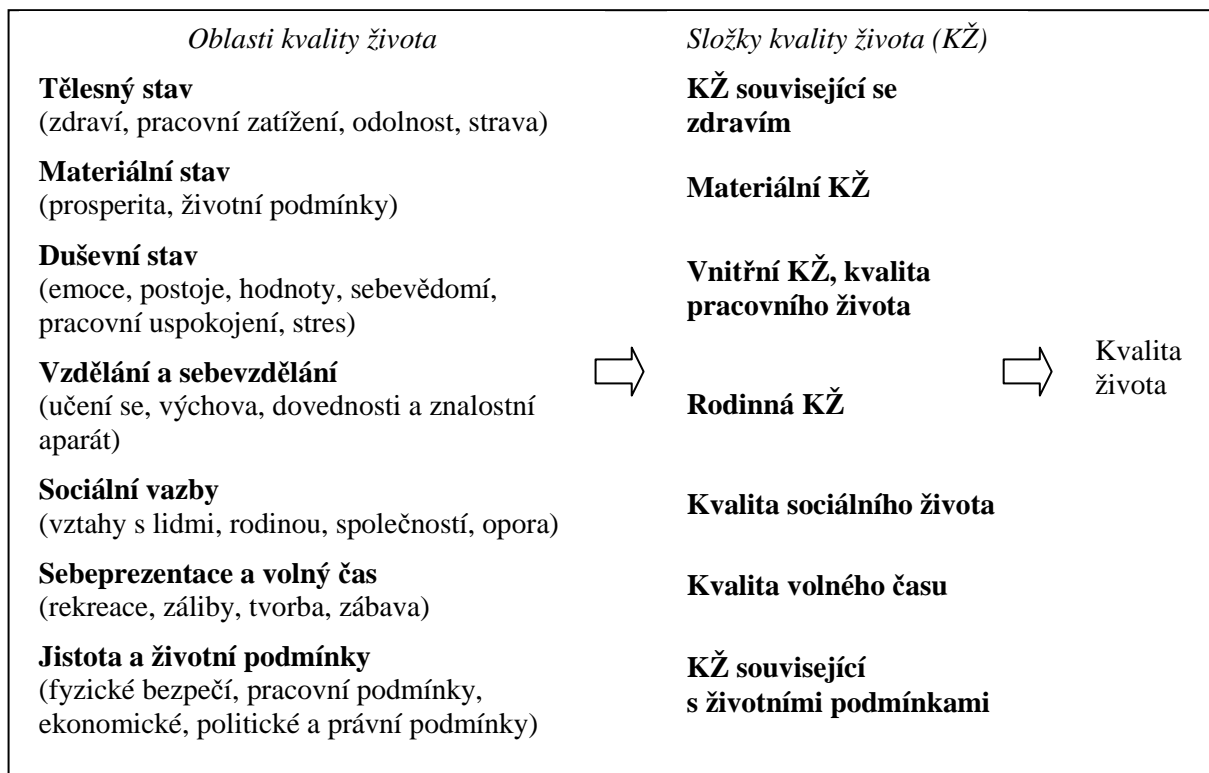
Obecně je možné říct, že kvalita života je ovlivněna individuálním fyzickým a duševním zdravím, stupněm nezávislosti, sociálním vztahem k životnímu prostředí a dalšími faktory. Kvalita života může být definována jako individuální spokojenost člověka s životem, který prožívá v porovnání s ideálním životem. Hodnocení kvality života závisí v tomto případě na hodnotovém systému každého člověka [1]. Kdyby však bylo nutné složitě se dotazovat každého jedince na jednotlivé aspekty jeho života, byl by tento postup časově velmi náročný a nebylo by možné provádět výzkum kvality života na neomezeném množství respondentů. Z tohoto důvodu se hledají jiné metody, které by čerpaly data z veřejně dostupných databází a které by poskytly relevantní výsledky.

Konceptuální modely kvality života a nástroje pro výzkum, posuzování a hodnocení byly vyvinuty ve druhé polovině minulého století. Nicméně už řeční filozofové hledali smysl života, který by mohl pomoci lidem najít vyšší existenciální úroveň jejich života. V minulém století byla kvalita života chápána jako materiální blahobyt nebo bohatství. Později došlo ke změně vnímání smyslu život a jeho hodnot, což ovlivnilo koncepci kvality života a všechny faktory se změnily.

Hlavním problémem je, že neexistuje žádné univerzální vymezení kvality života. V [1] je řečeno, že kvalita života závisí také na vnějších faktorech. Dobré životní podmínky nebo okolnosti určují vysokou kvalitu života, ale pokud se tyto podmínky liší, spokojenost jednotlivců s kvalitou života se liší také. Kvalita života je ovlivněna mnoha různými faktory a podmínkami jako jsou: ubytování, zaměstnanost, příjmy, materiální blahobyt, morální postoje, osobní a rodinný život, sociální podpora, stres a krize, zdraví, zdravotnictví, pracovní podmínky, výživa, vzdělávací příležitosti, ekologické faktory a další, jak vyplývá z obrázků 1 a 2. Originály použitých obrázků je možno nalézt v [1] nebo v příloze 1 a 2.



Obrázek 1 – Faktory které vytvářejí kvalitu života [upraveno podle 1]



Obrázek 2 – Model kvality života [upraveno podle 1]

Na kvalitu života je možné nahlížet ze dvou úhlů pohledu. Na jedné straně stojí subjektivní a na straně druhé objektivní kvalita života. Subjektivní kvalita života vychází z individuálně hodnocených podmínek života osoby, kam se promítají pozitivní i negativní emoce a rozdílné názory na svět. Naproti tomu objektivní kvalita života v sobě promítá podmínky života osoby nestranně, nezaujatě. Obecně lze říci, že se pojem kvality života pohybuje někde mezi životní úrovní a spokojeností, přičemž objektivnímu pólu je blíže v politice, ekonomii a subjektivnímu v medicíně, sociologii a psychologii [14].

Při studiu kvality života je třeba brát v úvahu rovněž vývoj, proměny v prostoru a čase, společenské souvislosti, historické a kulturní kořeny (západní křesťanská kultura, orientální kultura,...), civilizační i generační změny. Nelze opomenout ani často diskutované globální problémy a hrozby. Tyto bývají nejčastěji členěny do 3 skupin:

- intersociální problémy, které bývají spojovány se vztahem člověk-člověk. Řadí se sem problém války a míru a různé katastrofy, které mohou lidé eventuálně způsobit (biologické, jaderné a chemické), dále problém překonání sociálně ekonomické zaostalosti méně rozvinutých zemí, problém mezinárodní zadluženosti, dluhová krize rozvojových zemí a problém změn mezinárodních ekonomických vztahů, zpřístupnění trhů vyspělých zemí pro zboží z rozvojových států, spravedlivější pravidla pro obchod a oddlužení chudých zemí.

- přírodně sociální problémy, které mají počátek v porušení vazeb člověk-příroda. Do této skupiny se řadí populační problém, potravinový (nutriční) problém, surovinový a energetický problém, omezenost neobnovitelných zdrojů a ekologický problém - oslabování ozónové vrstvy, globální oteplování, skleníkový efekt a hrozba globálních klimatických změn, zvednutí hladin oceánů a moří apod.

- antroposociální problémy, které se týkají budoucnosti člověka, zahrnují všelidské problémy sociální, humanitární a kulturní. Příkladem problémů může být absolutní chudoba, drogové závislosti, šíření epidemií a dalších onemocnění, nekontrolovaná mezinárodní migrace či terorismus, živený politickými a národnostními konflikty.

V současné společnosti je ve vyspělých zemích zaznamenáváno posilování úlohy hmotných statků a odklon od hodnot nehmotných a relativně trvalých hodnot společnosti [14].

2.1 Indikátory kvality života

Indikátory kvality života mohou být nazvána jakákoliv číselná data, která mají anebo mohou mít nějaký vztah ke kvalitě života. Není známa žádná metoda, na základě které by bylo

možno se rozhodnout, které z možných ukazatelů vybrat pro další zpracování. Pro výzkum je možné použít proměnné použité jinými výzkumníky při studiu kvality života nebo dosud nepoužité proměnné, které splňují podmínku reprezentativnosti vzorků a mohou mít nějaký vztah ke kvalitě lidského života.

Sledováním a hodnocením indikátorů kvality života se už delší dobu zabývá více světových institucí, jako například Světová zdravotnická organizace, Evropská agentura životního prostředí a Eurostat. Díky jejich činnosti vznikly rozsáhlé soubory ukazatelů, které v členění na jednotlivé rozvojové oblasti sice přinášejí množství informací, avšak jsou málo integrované a pro většinu uživatelů neposkytují jednoduchý a souhrnný pohled na rozvojovou problematiku [10].

2.2 Index lidského rozvoje

Index lidského rozvoje (*Human development index*, HDI) je pokus o vyjádření kvality lidského života kombinací nejen ekonomických, ale i celé řady sociálních ukazatelů. Jako příklad je možno uvést očekávanou délku života, přístup ke vzdělání, rovné příležitosti z hlediska gender, porodní úmrtnost, přístup k nezávadné pitné vodě, zdravotní péči, a podobně. Vzhledem k typu ukazatelů, ze kterých je HDI počítán, je jeho využití omezeno na národní úroveň, pro nižší pozorovací úroveň nejsou dostupná data. Index je uveřejňovaný ve statistické publikaci „Zpráva o lidském rozvoji“, která je pod záštitou Rozvojového programu OSN (United Nations Development Programme). Používá se pro měření potenciální sociální prosperity. Index vymyslel v roce 1990 pákistánský ekonom Mahbub al Hak a od roku 1993 jej OSN používá ve své každoroční zprávě. Tento index rozděluje státy do čtyř skupin na základě jejich HDI, a to na země, jejichž HDI je: velmi vysoký (hodnota $> 0,900$), vysoký ($0,899-0,800$), střední ($0,799-0,500$) nebo nízký (hodnota indexu $< 0,499$) [9].

Na seznamu států světa podle indexu lidského rozvoje tak jak jej vydala OSN (aktualizace k 5.10.2009) figuruje 182 ze 192 členských států OSN, samostatně jsou uvedeny Hongkong (jako zvláštní administrativní území pod správou Číny) a okupovaná palestinská území. V seznamu chybí 10 členských států OSN, protože se nepodařilo získat potřebná data. Jsou mezi nimi například Irák, Libérie nebo Somálsko. Států s velmi vysokým indexem je uvedeno 38, s vysokým indexem 45 států, se střední hodnotou indexu 75 a s nízkou hodnotou indexu 24 států. Česká republika je v celkovém pořadí států na základě tohoto indexu uvedena na 36 místě [9][12]. Mapa světa s rozdělením HDI v členských státech OSN v roce 2007 je znázorněna v příloze 3.

3 Faktorová analýza

Matematických metod, které lze využít ke zkoumání každodenní reality, popisu jejích zákonitostí a celkového chování, je celá řada. Pro zpracování této diplomové práce byla vybrána faktorová analýza. Je to metoda, která v určitém smyslu vytváří hypotézy. Vychází z toho, že spolu více měřených proměnných úzce souvisí a jsou vzájemně silně korelované. Dá se pak říci, že tyto proměnné vyjadřují v určité míře to samé a dá se předpokládat, že se navzájem určují, anebo že se v nich projevuje třetí veličina, kterou je možné přímo měřit. Tyto závislosti se pokouší zachytit model faktorové analýzy zobrazený na obrázku 3. Tento předpoklad je v mnohých případech realistický. Použitím faktorové analýzy je zjišťováno, zda se z proměnných, které byly pozorovány, dá izolovat veličina, tzv. faktor, který by objasnil pozorované souvislosti. Slovo faktor je v tomto případě používáno ve smyslu matematické veličiny odvozené z pozorování. Na základě této definice se lze ptát, kolik a jakých hypotetických veličin je nutné stanovit na to, aby bylo možno reprodukovat a vysvětlit pozorované vztahy mezi proměnnými co nejpřesněji nebo jak se dají zredukovat větší soubory dat na co nejjednodušší pojmy. Faktory mají být co nejjednodušší a mají dostatečně přesně popisovat a vysvětlovat pozorování [2].

Faktorová analýza patří do skupiny statistických modelů označovaných jako modely s latentními proměnnými, je nejznámějším z těchto modelů. Jako modely s latentními proměnnými se rozumí takové modely, které „popisují a v jistém smyslu vysvětlují pozorovaná data pomocí jejich závislosti na nepozorované charakteristice, kterou lze za určitých předpokladů matematicky zkonstruovat [2].“

Aplikace modelů s latentními proměnnými byly zpočátku rozšířené především v psychologii, sociologii, genetice a pedagogice, později se jich začalo využívat i v jiných oborech lidské činnosti, např. v medicíně, geologii, meteorologii, biologii a archeologii.

Modely s latentními proměnnými mohou být použity s různým účelem. Při aplikaci lze všeobecně rozlišit následující účely či typy použití [2]:

- redukce počtu proměnných, zjednodušení popisu jevů,
- klasifikace proměnných (taxonomie), obvykle spojená s vytvářením obecnějších druhových proměnných aglomerováním původních proměnných,
- odhad a zprostředkující měření nepřímo měřitelných proměnných,
- tvorba anebo ověření nějaké strukturální teorie o zkoumané oblasti,
- transformace původních proměnných do výhodnějšího tvaru.

Uvedené účely použití přicházejí v úvahu prakticky ve všech oborech, kde je analýza aplikovaná.

Základní myšlenky faktorové analýzy byly diskutovány přibližně od 2. poloviny 19. století. Za vlastní počátek faktorové analýzy se všeobecně považuje publikace anglického psychologa Charlese Spearmana, která vyšla roku 1904 pod názvem *General intelligence objectively determined and measured*. Spearman vycházel z toho, že jistý všeobecný faktor, který označil *g*, rozhodujícím způsobem ovlivňuje všechny intelektuální výkony. Představa jednoho společného faktoru však byla neudržitelná a vývoj vedl k tzv. vícenásobné faktorové analýze. Vícenásobná faktorová analýza umožňuje rozklad korelační matice na překrývající se faktory. Další vývoj faktorové analýzy byl poznamenán především vzrůstajícím vlivem matematické statistiky a elektronických počítačích strojů. Použití elektronických počítačích strojů mělo zpočátku pouze urychlit a ulehčit práci. Ulehčení práce bylo dosti významné, vznikla nová situace, která umožnila vykonávat výpočet řešení metodou hlavních os (metoda hlavních os byla matematicky formulována Pearsonem r. 1901) a všechny novější a náročnější extrakční metody a rotace [15].

Jsou známy dva přístupy k faktorové analýze. Prvním z nich je exploratorní přístup. Explorativní faktorová analýza představuje tradiční přístup k faktorové analýze, je její nejčastější formou. Snaží se odhalit základní strukturu relativně velkého souboru proměnných na základě zjištěné korelační matice. Dodatečně se pak hledá věcná interpretace faktorů jako nějakých obecnějších proměnných v daném aplikačním oboru. Tento postup se nazývá *ex post facto*, je vhodný k dodatečné tvorbě strukturální teorie pro daný obor. Druhým přístupem je konfirmatorní přístup, který tvoří protějšek tradiční exploratorní faktorové analýzy. Principem tohoto přístupu je, že se nejprve vytvoří hypotéza, která se následně experimentálně ověřuje. Podle hypotézy se vytvoří struktura očekávané matice faktorových zátěží a ta se prokládá do zadané kovarianční či korelační matice. Tato hypotéza o uspořádání, velikosti a nulových hodnotách koeficientů a_{jp} v matici zátěží A , se statisticky testuje, postupně upravuje a výsledkem je matice zátěží vystihující strukturu korelací bez tradičních metod rotace [2].

3.1 Základní pojmy

Má-li být v práci pracováno s faktorovou analýzou, je důležité definovat některé důležité pojmy, které se v souvislosti s ní používají, a také základní značení.

Manifestní proměnné

Proměnné, které jsou exaktně zjišťované. Značí se Y_j , $j = 1, 2, \dots, n$, kde n je počet proměnných. Např. projevy psychické (reakce v určité situaci, výsledky v testech inteligence,...), v lékařství jednotlivé příznaky nemoci (zvýšená teplota, zvýšený tlak,...), ve sportu sportovní výkony (délka skoku, rychlost běžce,...) apod.

Latentní proměnné

Soustavě F_p latentních proměnných, kde $p = 1, 2, \dots, m$, lze přisoudit věcný smysl, např. v medicíně jsou to syndromy nemoci, v psychologii se může mluvit o rysech osobnosti, ve sportu vytrvalost či kondice apod.

Cílem modelu latentních proměnných je podat popis (a v jistém smyslu vysvětlení) manifestních proměnných a jejich vzájemných vztahů za pomoci latentních proměnných.

Nevysvětlená část

Je možno použít také označení chyba měření nebo chyba testování, nese označení ε . V teorii chyb se tradičně používá model, který lze zapsat ve tvaru

$$y = a + \varepsilon \quad (1)$$

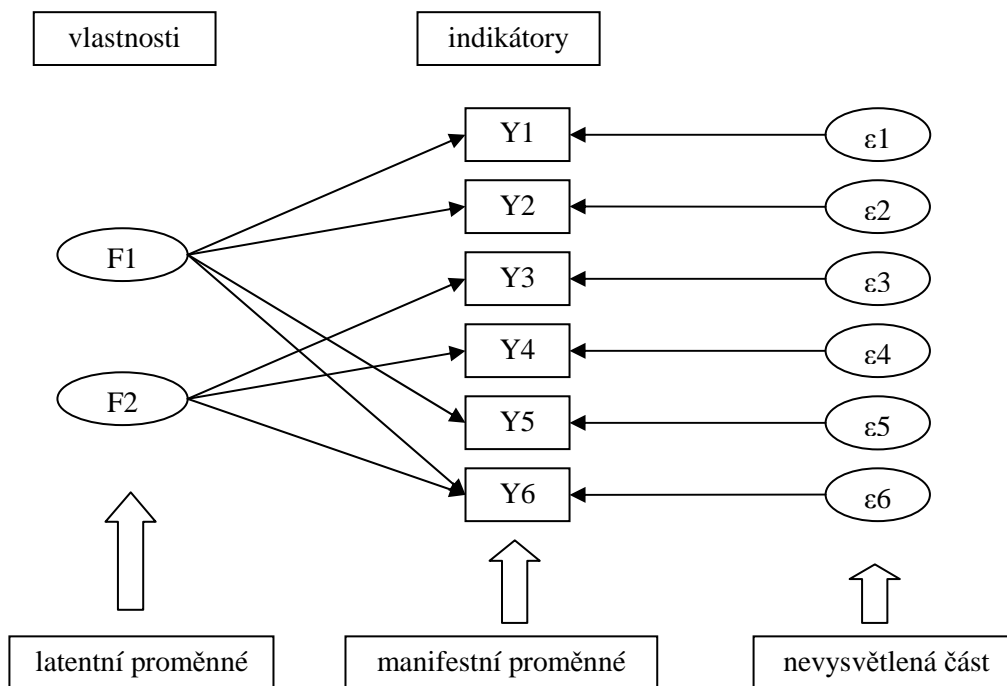
kde: y je naměřená hodnota,

a přesná hodnota,

ε je chyba měření.

Faktorový vzor, v této práci označovaný A , je tzv. maticí faktorových zátěží či faktorových koeficientů (ang. factor pattern). Jedná se o matici regresních koeficientů a_{jp} pozorovaných proměnných y_j na latentních faktorech f_p . V případě, že faktory f_p jsou navzájem nekorelované (ortogonální), jsou a_{jp} shodné s korelačními koeficienty $r(y_j, f_p) = q_{jp}$. Tvoří matici $Q = [q_{jp}]$ a nazývají se koeficienty faktorové struktury. Pro ortogonální faktory platí $A = Q$.

Rozptyl pozorovaných hodnot lze rozdělit na komunalitu, specificitu a rozptyl chyb. **Komunalita** proměnné Y_j udává část jejího rozptylu, která je vysvětlena působením společných faktorů. **Specificita** je rozptyl specifických částí výsledků. Rozptyl skutečných hodnot tvoří komunalita a specificita dohromady. Specificita a rozptyl chyb zase tvoří složku rozptylu označovanou jako jedinečnost (ang. uniqueness). **Jedinečnost** je tedy ta část rozptylu změřené manifestní proměnné, která je pro ni jedinečná v důsledku jejích specifických vlastností nebo chyb měření [2].



Obrázek 3 – Model faktorové analýzy, [zdroj: 13]

Základní myšlenka faktorové analýzy spočívá v tom, že vzájemná lineární závislost pozorovaných proměnných Y_1, Y_2, \dots, Y_n se vysvětluje existencí menšího počtu nepozorovaných (a nepozorovatelných) veličin F_1, F_2, \dots, F_m ($m < n$) (tzv. *společné faktory* či jenom *faktory*) a dalších n zdrojů variability $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ (tzv. *chybové* či *specifické faktory*).

Společné faktory vyvolávají korelace mezi proměnnými, zatímco chybové faktory pouze přispívají k rozptýlení jednotlivých pozorovaných proměnných. Model tedy lze zapsat lineárními rovnicemi (2) jako

$$\begin{aligned}
 Y_1 &= a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + \varepsilon_1 \\
 Y_2 &= a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + \varepsilon_2 \\
 &\vdots \\
 Y_n &= a_{n1}F_1 + a_{n2}F_2 + \dots + a_{nm}F_m + \varepsilon_n
 \end{aligned}
 \tag{2}$$

kde Y_n jsou měřené, manifestní proměnné,

a_{nm} faktorové zátěže, převodní koeficienty F_m na Y_n ,

F_m neznámé, neměřené latentní příčiny společně ovlivňující Y_n způsobem vyjádřeným rovnicemi,

ε_n chyba rovnice

nebo vektorově (3) jako

$$y = Af + \varepsilon \quad (3)$$

kde A je matice faktorových zátěží typu $n \times m$,

$f = (f_1, f_2, \dots, f_m)'$ je vektor faktorů,

$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ je vektor chybových složek [4].

3.2 Předpoklady faktorové analýzy

Jako v podstatě každá metoda analýzy dat statistického charakteru naráží také faktorová analýza na nejrůznější omezení, která se týkají charakteru zkoumaných proměnných a vztahů mezi nimi. Zde jsou uvedena pouze ty nejdůležitější:

1. Dostatečná velikost vzorku – minimální doporučené množství případů ve výběru je rovno počtu extrahovaných faktorů.
2. Jsou předpokládána intervalová data. Ordinální data mohou být za určitých podmínek použita také, ale výsledkem jejich užití může být obtížnější interpretace faktorů. Obdobně je možné využití dichotomických údajů v případě, že základní metrické korelace mezi proměnnými jsou považované za mírné (0.7) nebo nižší.
3. Důležitým předpokladem faktorové analýzy je, aby vzájemné korelace jednotlivých závisle proměnných byly dostatečně vysoké. Často bývá požadováno, aby byly tyto korelace větší než 0,3.
4. Předpokladem smysluplného výstupu faktorové analýzy je rovněž linearita zkoumaných vztahů, neboť faktorová analýza je v principu lineární procedura.
5. Na předpoklad linearity vztahů navazuje předpoklad homoskedasticity, nicméně ani tento předpoklad není považován za příliš podstatný.
6. Stejně tak není rozhodující požadavek vícerozměrného normálního rozložení, které je sice vyžadováno pro řadu statistických testů, které faktorovou analýzu doprovázejí, ovšem samotná procedura dodržování normality nevyžaduje. Čím menší je velikost vzorku, tím důležitější je, aby zkoumaná data měla normální rozložení pravděpodobnosti.
7. Odlehlá pozorování mohou zkreslit výsledek faktorové analýzy, proto je důležité tyto případy odhalit a eliminovat jejich vliv.
8. Multikolinearita – přítomnost této vlastnosti analyzovaného datového souboru indikuje přítomnost redundantní proměnné, která vede k tomu, že vstupní matice je singulární, tudíž nelze spočítat matici k ní inverzní, která je potřebná k provádění výpočtů v rámci faktorové analýzy. Je tedy předpokládána absence multikolinearity [5][6].

3.3 Rotace

Rotace ve faktorové analýze se používá pro větší srozumitelnost výstupu a jeho snazší interpretovatelnost. Při faktorové rotaci jde o snahu, aby každá proměnná měla vysoké faktorové zátěže u co nejmenšího počtu společných faktorů a zátěže blízké nule u ostatních faktorů. Řešení bez použití rotace bývá obtížné objasnit právě proto, že proměnné mívají faktorové zátěže u více než jednoho faktoru. Rozlišují se přitom dva druhy rotace, přičemž u každého z nich lze uvést více metod řešení problému.

1. Ortogonální (pravoúhlá) rotace – poskytuje řešení s nekorelovanými faktory

Varimax je nejrozšířenější metoda rotace. Každý faktor zde má tendenci mít buď velké, nebo malé zatížení ve vztahu ke každé jednotlivé proměnné. Řešení je snadné interpretovat právě proto, že lze jednoznačně přiřadit každou proměnnou k jednomu faktoru.

Quartimax minimalizuje počet faktorů potřebných k vysvětlení proměnných. Tento typ rotace často vytváří „generální faktor“, k němuž se vztahují vysoké nebo střední faktorové zátěže proměnných. Tato struktura faktorů obvykle není užitečná pro účely zkoumání.

Equamax je možné nazvat střením cestou mezi Varimax a Quartimax rotací.

2. Šikmá rotace - výsledné faktory jsou korelované

Direct oblmin je standardní metodou šikmé rotace. Následkem šikmé rotace se zvýší vlastní čísla, ale sníží se interpretovatelnost výstupu.

Promax je obdobou předchozího, ale na rozdíl od direct oblmin je výpočetně rychlejší, a proto se používá pro velmi rozsáhlé soubory dat [3][5].

3.4 Metody získávání faktorů

Metod extrahování faktorů existuje celá řada. Zde je uvedeno šest z nich.

Analýza hlavních komponent (PCA) - jedná se o základní a nejběžnější metodu faktorové analýzy, používá se především při předběžném výzkumu, kdy je potřeba snížit vysoký počet proměnných na menší počet latentních proměnných. Cílem metody je, aby každý faktor reprezentoval maximum rozptylu [5].

Metoda nevážených nejmenších čtverců je založená na minimalizaci součtu kvadratických rozdílů mezi pozorovanou a odhadnutou korelační maticí.

Zobecněná metoda nejmenších čtverců je podobná metodě nevážených nejmenších čtverců, ale zde je použito vážení. Korelace je vážena nepřímo úměrně na základě jedinečnosti (čím větší jedinečnost, tím nižší váha) [3].

Metoda maximální věrohodnosti je statistickou metodou, jejímž výsledkem jsou takové odhadnuté parametry základního souboru, které mají co největší pravděpodobnost, že se vyskytnou u výběru z tohoto základního souboru. Metoda se snaží usuzovat z výběru na základní soubor, je založena na předpokladu souvislosti. Předpokládá se, že pozorované proměnné mají normální rozložení a že společné faktory mají být ortogonální [15].

Metoda hlavních os je metodou, která je velmi podobná metodě hlavních komponent, zaměřuje se na společné odchylky [3].

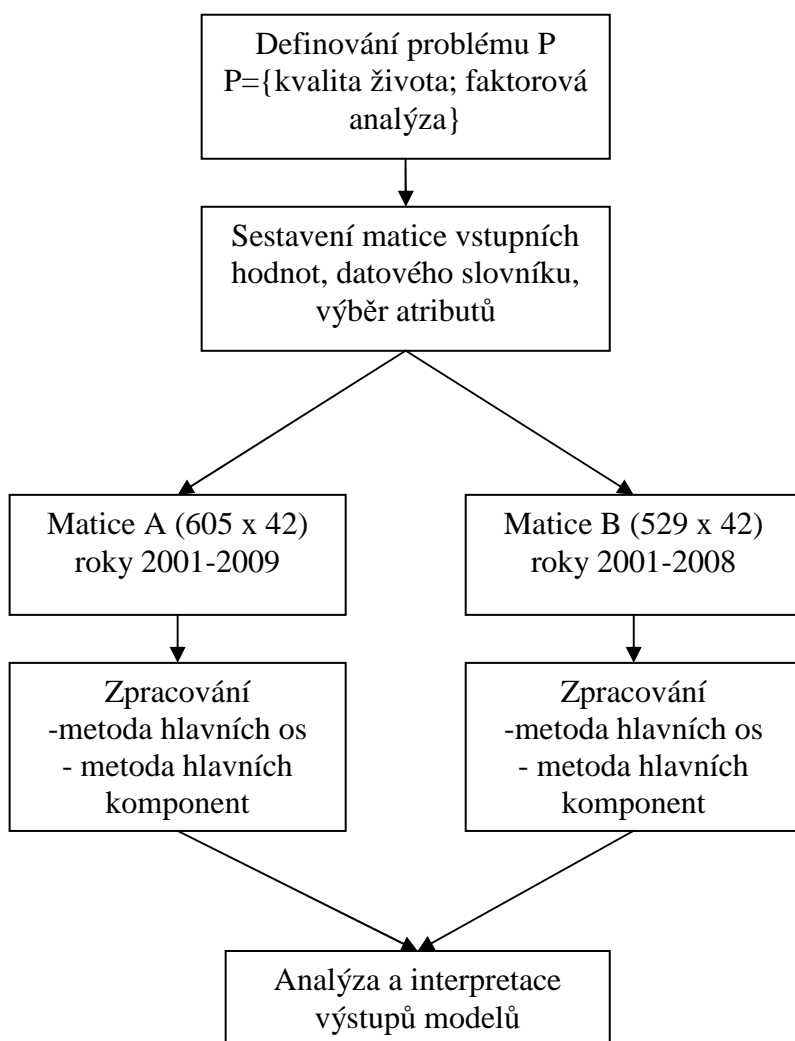
Alfa faktoring, tato metoda faktorové analýzy využívá koncept koeficientu α , čímž se rozumí korelační matice pozorované proměnné se skutečnou proměnnou. Extrahuje takové faktory, které vykazují maximální korelaci s odpovídajícími faktory základního souboru proměnných. Extrahují se všechny faktory s vlastními hodnotami většími než 1, protože těmto faktorům odpovídá pozitivní koeficient α [15].

V praxi bývá často zaměňována faktorová analýza s metodou hlavních komponent. Obě tyto metody patří do metod vícerozměrné statistické analýzy. Vícerozměrná statistická analýza je poměrně mladá disciplína, jejíž rozvoj a uplatnění v praxi bylo podmíněno vývojem výpočetní techniky a programového vybavení. Teoretické základy byly položeny ve 30. a 40. letech 20. století [8]. Z porovnání faktorové analýzy a metody hlavních komponent (PCA) plyne následující:

1. Ani jednu metodu nemá cenu použít, pokud jsou původní proměnné nekorelované. Faktorová analýza pak nemá co objasnit a PCA vede k hlavním komponentám identickým s původními proměnnými.
2. Faktorová analýza postuluje model pro data, PCA nikoliv.
3. Faktorová analýza se pokouší vysvětlit kovariance a korelace původních proměnných pomocí několika málo společných faktorů. PCA objasňuje pouze rozptyl původních proměnných.
4. Pokud bude u PCA zvýšen počet použitých proměnných A o 1 na $A+1$, původní komponenty se nezmění. Pokud přidáme ve faktorové analýze další faktor, ostatní faktory se podstatně změní.
5. Výpočet PCA je přímočarý, jednoduchý, zatímco výpočet faktorového skóre u faktorové analýzy je daleko komplexnější a byla pro něj navržena řada postupů.
6. Obyčejně není žádný vztah mezi hlavními komponentami a korelační maticí nebo kovarianční maticí, vztah mezi faktory a korelační či kovarianční maticí je vždy [11].

4 Formulování problému

V zadání práce je uvedeno, že bude provedeno modelování kvality života na základě faktorové analýzy. Práce vychází z předpokladu, že pro modelování v dané oblasti lze tuto metodu použít. Prvním krokem vedoucím k vypracování zadání je, jak je vidět z obrázku 4, formulování problému P, který sestává z prvků kvalita života a faktorová analýza. V této části je stanoven postup vedoucí k naplnění cíle. Rámcové seznámení s pojmem kvality života i faktorové analýzy bylo provedeno již v předchozí části práce. Pro zpracování byl zvolen exploratorní přístup k faktorové analýze, protože dosud nebyl zjištěn žádný předpoklad, na jehož základě by se vytvořila struktura očekávané matice faktorových zátěží potřebná pro konfirmatorní přístup. Po seznámení s problematikou kvality života z obsahového hlediska a po seznámení se současnou situací ve světě se práce dále věnuje kvalitě života v České republice. Rozlišovací úrovní pro sběr dat byly stanoveny okresy.



Obrázek 4 – Navržený model řešení [zdroj: vlastní]

V dalším kroku byla sestavena datová matice. Při rozhodování, které proměnné budou do datové matice zařazeny, bylo vycházeno z teoretických základů kvality života a z dostupnosti dat. Jedním ze stanovených předpokladů bylo, že data použitá pro modelování budou získatelná z veřejně dostupných databází. Základním zdrojem dat je v České republice Český statistický úřad a jeho krajské pobočky, odkud byla získána i převážná část dat použitých v této diplomové práci. Chybějící údaje byly doplněny ze statistik zveřejněných na internetových stránkách risy.cz.

Protože nebyl nalezen žádný seznam proměnných, které ovlivňují kvalitu života, byly na základě přístupů ke kvalitě života vybírány takové, které podle názoru autorky mají anebo mohou mít nějaký vztah ke kvalitě lidského života. Počet vstupních proměnných modelu byl dále omezen již zmíněnou dostupností dat, protože na úrovni okresů není v České republice mnoho proměnných sledováno vůbec a mnohé pouze v některých letech. Byl stanoven předběžný seznam proměnných, ze kterých bude vytvořena matice vstupních hodnot pro faktorovou analýzu. Následovala časově velmi náročná fáze, a to vlastní sběr dat a jejich uspořádání do matice. Proměnné, u kterých nebyly nalezeny data ve všech sledovaných letech, tj. v letech 2001-2009*, byly vyjmuty z výsledné datové matice. Jak již bylo zmíněno v předpokladech faktorové analýzy, manifestní proměnné vstupující do faktorové analýzy musí být korelované. V případě, že by korelované nebyly, vytvořila by faktorová analýza tolik faktorů, kolik je vstupních proměnných. V práci není korelovanost pozorovaných proměnných ověřována, předpokládá se, že proměnné korelované jsou.

4.1 Datová matice

Základem pro modelování je vytvoření datové matice, kde jsou jednotlivé záznamy reprezentovány řádky a pozorované proměnné sloupci. Záznamy jsou zde myšleny okresy České republiky ve sledovaných letech. Pro všech 76 okresů by tedy mělo existovat 9 záznamů, pro každý rok jeden. Tento fakt však v práci není naplněn, protože pro některé okresy nebyla data nalezena. Týká se to okresů Plzeňského kraje (7 okresů) v letech 2001 a 2002, okresů Královéhradeckého kraje (5 okresů) v letech 2001 – 2004, okresů kraje Vysočina (5 okresů) v letech 2001 – 2003, okresů Jihomoravského kraje (7 okresů) v letech 2001 a 2002 a okresů Zlínského kraje (4 okresy) v letech 2001 – 2004. Výsledný počet záznamů v datové matici je tedy 605. Počet pozorovaných proměnných se ustálil na číslovce 42. Problém se vyskytl v posledním sledovaném roce, kde nebylo 11 proměnných vykazováno. Vzhledem k tomu, že není žádoucí vzdát se jedenácti proměnných bez zjištění

* Uvedené roky jsou převzaty z ročníku statistických ročenek, nejedná se o rok, za který jsou data sledována.

jejich vlivu na výsledný model, budou vytvořeny 2 modely. Jeden pro datovou matici označenou A, kterou je možné matematicky zapsat

$$y_j^h(y_1, y_2, \dots, y_{42}), \quad (4)$$

kde y jsou pozorované proměnné, $y = 1, 2, \dots, 42$

h počet okresů, $h = 76$

j roky, $j = 2001, 2002, \dots, 2009$.

Hodnoty jedenácti proměnných, které nejsou v roce 2009 vykazovány, jsou doplněny pomocí funkce medián, což je číslo, které leží uprostřed podle velikosti uspořádaného souboru čísel. Polovina čísel má tedy hodnotu, která je větší nebo rovna mediánu a polovina čísel má hodnotu, která je menší nebo rovna mediánu. Do funkce vstupují vždy hodnoty za poslední čtyři roky, výsledné číslo je dáno průměrem ze dvou prostředních. Interval čtyř let byl vybrán z důvodu aktuálnosti dat (vzhledem k věcnému významu proměnných jsou starší data pro doplnění chybějících hodnot nevhodná) a také proto, aby byl časový úsek pro zjišťování mediánu u všech okresů shodný. Matice má rozměr 605 x 42 polí, což dává dohromady 25410 údajů.

Matematický zápis druhého modelu, jehož vstupní matice je označena B, je totožný se zápisem matice A v rovnici 4. Liší se pouze v parametru j , který na rozdíl od předchozího modelu, obsahuje hodnoty $j = 2001, 2002, \dots, 2008$. Velikost výsledné matice je tedy 529 x 42, tj. 22218 dat.

Pro index h platí, že pro oba modely nejsou uvedeny záznamy pro výše uvedené okresy jednotlivých krajů.

Poznámka: Jednotlivé roky jsou uváděny podle ročníku statistických ročenek, tzn. že, se nejedná se o rok, za který jsou data sledována. Údaje jsou o rok starší, tj. v roce 2001 jsou vykazovány údaje za rok 2000 apod. Výjimka z uvedeného platí pro proměnné s kódovým označením P11, P12, P13 a P14. Emise jsou uváděny s ročním zpožděním, tj. např. v ročence 2007 jsou uvedeny hodnoty emisí za rok 2005.

4.2 Datový slovník

Datový slovník obsahuje základní charakteristiku zpracovávaných dat. Proměnné jsou po celou dobu práce s nimi, tj. od sběru dat, přes práci s datovými maticemi až po modelování, prezentovány pouze pod kódovým označením. V datovém slovníku je stručně popsána každá proměnná, její měrná jednotka a základní charakteristika dat, kam se řadí

aritmetický průměr, minimální hodnota, maximální hodnota a rozsah, tj. maximální hodnota-minimální hodnota. Všechny údaje jsou přehledně zpracovány v tabulce 1.

V buňkách náležejících základním charakteristikám dat je uvedena jedna nebo dvě hodnoty. Pokud jsou hodnoty základních charakteristik shodné pro matici A i B, je uvedena pouze jedna hodnota. V případě že se liší, náleží první údaj k matici A, druhý k matici B.

Tabulka 1 – Datový slovník [zdroj: vlastní]

Kód	Název proměnné, popis obsahu pole				Měrná jednotka
	Základní charakteristiky dat				
	Aritmetický průměr	Minimální hodnota	Maximální hodnota	Rozsah	
P1	Rozloha okresu celkem				km ²
	1025,5; 1024,7	124,8	1946,0	1821,2	
P2	Hustota obyvatelstva				osoby/km ²
	172,4; 173,6	36,8	1610,0; 1609,5	1573,1; 1572,6	
P3	Průměrný věk obyvatel				roky
	39,4; 39,3	36,2	41,9	5,7	
P4	Podíl vybraných druhů pozemků - zemědělská půda z celkové rozlohy okresu				%
	53,5	27,2	74,7	47,4	
P5	Dokončené byty, místnost nebo soubor místností, které mohou sloužit jako samostatné bytové jednotky, na které ve sledovaném období nabyla právní moci vydaná kolaudační rozhodnutí				na 1000 obyvatel
	3,0	0,3	25,6	25,3	
P6	Hromadná ubytovací zařízení				celkem
	103,6; 105,2	15,0	796,0	781,0	
P7	Průměrný měsíční důchod celkem				Kč
	7498,8; 7238,7	5872,0	9680,0; 8824,0	3808,0; 2952,0	
P8	Průměrný měsíční důchod starobní				Kč
	7641,8; 7374,4	6054,0	9984,0; 9119,0	3930,0; 3065,0	

Kód	Název proměnné, popis obsahu pole				Měrná jednotka
	Základní charakteristiky dat				
	Aritmetický průměr	Minimální hodnota	Maximální hodnota	Rozsah	
P9	<p>Maloplošná chráněná území, do této kategorie se řadí 4 typy území: <i>národní přírodní památky</i> - přírodní útvar menší rozlohy, naleziště nerostů či ohrožených druhů ve fragmentech ekosystémů s národním či mezinárodním významem ekologickým, vědeckým či estetickým <i>přírodní památky</i> - obdobně, ale pouze s regionálním významem <i>národní přírodní rezervace</i> - menší území mimořádných přírodních hodnot, kde jsou na přirozený reliéf s typickou geologickou stavbou vázány ekosystémy významné v národním či mezinárodním měřítku <i>přírodní rezervace</i> - menší území soustředěných přírodních hodnot se zastoupením ekosystémů typických a významných pro příslušnou geografickou oblast [ČSÚ]</p>				ha
	1174,1; 1177,8	58,0	6814,0	6756,0	
P10	Pořízené investice na ochranu životního prostředí podle okresu sídla investora, samostatné stavby a ostatní investiční opatření, vedoucí ke zlepšení současného stavu životního prostředí				v tis. Kč běžných cen
	225813,5; 226137,3	8878,0	4280993,0	4272115,0	
P11	Emise základních znečišťujících látek do ovzduší - emise tuhé; REZZO 1-3 (velké střední a malé zdroje znečišťování)				kg/obyvatele
	5,3; 5,5	0,3	15,7	15,4	
P12	Emise základních znečišťujících látek do ovzduší - oxid siřičitý (SO ₂); REZZO 1-3				kg/obyvatele
	26,3; 26,8	0,4	343,2	342,8	
P13	Emise základních znečišťujících látek do ovzduší - oxidy dusíku (NO _x); REZZO 1-3				kg/obyvatele
	18,1; 18,3	1,5	215,5	214,0	
P14	Emise základních znečišťujících látek do ovzduší - oxid uhelnatý (CO); REZZO 1-3				kg/obyvatele
	25,7; 26,4	0,9	360,4; 274,6	359,5; 273,7	

Kód	Název proměnné, popis obsahu pole				Měrná jednotka
	Základní charakteristiky dat				
	Aritmetický průměr	Minimální hodnota	Maximální hodnota	Rozsah	
P15	Sňatky, jsou započítávány do příslušného území podle místa trvalého bydliště ženicha				na 1 000 obyvatel
	5,0	-0,4	6,4	6,8	
P16	Rozvody, jsou územně tříděny podle posledního společného bydliště manželů				na 1 000 obyvatel
	3,1	1,9	6,1	4,2	
P17	Potraty, tj. předčasné ukončení těhotenství, které lékař klasifikuje jako potrat.				na 100 narozených
	42,4; 43,4	21,0; 24,0	83,6	62,6; 59,6	
P18	Celkový přírůstek obyvatelstva, vyjadřuje rozdíl mezi počátečním stavem obyvatelstva a koncovým stavem obyvatelstva daného území, je tvořen součtem přirozeného přírůstku a přírůstku stěhováním.				na 1 000 obyvatel
	3,1; 2,5	-12,3	64,3; 55,5	76,6; 67,9	
P19	Míra registrované nezaměstnanosti, vyjadřuje podíl nezaměstnaných (čitatel) na celkové pracovní síle (jmenovatel). Šetřené osoby jsou uváděny podle místa jejich pobytu.				%
	8,8; 9,2	1,6	23,5	21,9	
P20	Volná pracovní místa				celkem
	737,1; 717,8	44,0	8550,0	8506,0	
P21	Délka silnic a dálnic				km
	728,6; 728,2	105,0	1245,0	1140,0	
P22	Lékaři				na 1 000 obyvatel
	3,3	1,5	8,5	6,9	
P23	Počet obyvatel na 1 lékaře				přepočtené osoby
	321,5; 323,6	117,9	651,0	533,1	

Kód	Název proměnné, popis obsahu pole				Měrná jednotka
	Základní charakteristiky dat				
	Aritmetický průměr	Minimální hodnota	Maximální hodnota	Rozsah	
P24	Nemocnice				celkem
	2,2; 2,3	0,0	12,0	12,0	
P25	Lékárny a výdejny léků (lékárny včetně odloučených pracovišť)				celkem
	26,3; 25,9	7,0	122,0	115,0	
P26	Průměrný počet nemocensky pojištěných				osoby
	46767,0; 46733,7	10950,0; 11302,0	279594,0; 275580,0	268644,0; 264278,0	
P27	Průměrná doba trvání 1 případu pracovní neschopnosti				kalendářní dny
	33,6; 32,7	23,1	49,4; 44,0	26,2; 20,9	
P28	Průměrné procento pracovní neschopnosti za rok, vypočítá se jako podíl počtu kalendářních dnů pracovní neschopnosti pro nemoc a úraz a průměrného počtu zaměstnanců nemocensky pojištěných, násobeného počtem kalendářních dnů v roce.				%
	6,4; 6,5	4,3; 4,6	9,5	5,2; 4,9	
P29	Zjištěné trestné činy, ukazatel zahrnuje obecnou i hospodářskou kriminalitu. Obecná kriminalita zahrnuje násilné, mravnostní a majetkové trestné činy, podvody a zpronevěry, ostatní majetkové a ostatní kriminální trestné činy, zatímco hospodářská kriminalita zahrnuje především trestné činy v obchodních a finančních vztazích.				na 1 000 obyvatel
	28,1; 28,2	12,8	52,2	42,4	
P30	Dopravní nehody, zahrnují všechny nehody nahlášené Policii České republiky.				na 1 000 obyvatel
	17,4; 17,9	7,5; 7,8	39,3	31,7; 31,4	
P31	Požáry				celkem
	254,5; 256,2	51,0	869,0	818,0	

Kód	Název proměnné, popis obsahu pole				Měrná jednotka
	Základní charakteristiky dat				
	Aritmetický průměr	Minimální hodnota	Maximální hodnota	Rozsah	
P32	Stálá kina, v případě, že je v jedné budově provozováno více kinosálů (např. multikino), je toto zařízení považováno pouze za jedno stálé kino.				celkem
	7,9; 8,0	1,0	21,0	20,0	
P33	Veřejné knihovny včetně poboček, údaje jsou vedeny za všechny typy knihoven (okresní, místní lidové a ostatní) s profesionálními pracovníky				celkem
	84,4; 84,5	25,0	175,0	150,0	
P34	Muzea				celkem
	9,8; 9,7	2,0	23,0	21,0	
P35	Galerie				celkem
	10,4; 10,2	0,0	57,0	57,0	
P36	Divadla, ukazatel je vztažen k budovám. Nejsou započítány divadelní budovy a sály, které slouží jen jako zájezdové scény.				celkem
	2,4	0,0	15,0	15,0	
P37	Přírodní amfiteátry, víceúčelová otevřená zařízení určená k letnímu pořádání divadelních a filmových představení, estrád a dalších kulturně společenských akcí. Tento ukazatel zahrnuje také letní kina.				celkem
	4,3; 4,2	0,0	15,0	15,0	
P38	Koupaliště a bazény, koupaliště jsou zařízení u vodních toků a nádrží, která mají provozovatele. Pokud je na jedné ploše vybudováno více koupališť, pak je každé z nich bráno jako samostatné zařízení.				celkem
	16,1	1,0	97,0	96,0	
P39	Hřiště				celkem
	125,9; 125,5	16,0	367,0	351,0	

Kód	Název proměnné, popis obsahu pole				Měrná jednotka
	Základní charakteristiky dat				
	Aritmetický průměr	Minimální hodnota	Maximální hodnota	Rozsah	
P40	Tělocvičny				celkem
	65,4; 65,2	21,0	147,0	126,0	
P41	Stadiony včetně krytých				celkem
	11,9	2,0	33,0	31,0	
P42	Zimní stadiony včetně krytých				celkem
	2,7	0,0	9,0	9,0	

Ukázka prvních patnácti záznamů datové matice je zobrazena v příloze 4.

4.3 Zpracování

Po seznámení se se vstupními daty a jejich základními charakteristikami již nic nestojí v cestě vytvoření modelu. Pro modelování byl použit produkt od společnosti SPSS Clementine 10.1. Jedná se o software, který je vhodný k analýze dat a vytváření modelů od neuronových sítí přes rozhodovací stromy, shlukovou analýzu a různé druhy regrese. Nechybí zde ani možnost faktorové analýzy, pro kterou se využívá uzel PCA/Factor. Defaultním nastavením tohoto uzlu je pro extrakci faktorů metoda hlavních komponent (Principal Components), využití korelační matice pro tvorbu faktorů. Maximální počet iterací konvergence je stanoven na 25, hodnota vlastních čísel faktorů větší než 1 a je určeno, že nebude použita žádná metoda rotace. Počet iterací nebyl v průběhu modelování měněn. Počet iterací výstupních modelů byl vždy nižší než zadaný počet. Oproti implicitnímu nastavení, kdy není používána rotace, byla při tvorbě každého modelu zvolena rotační metoda Varimax.

Při modelování byly použity dvě metody pro extrakci faktorů: metoda hlavních os a metoda hlavních komponent. Pomocí experimentů bylo vytvořeno několik modelů a každý z nich popsán. Nejdřív byla vstupem modelu matice A o rozměrech 605 x 42 polí, posléze matice B o rozměru 529 x 42 polí. Z popsaného vyplývá, že byly vytvořeny čtyři modely postupně označované A, B, C a D.

Pro každou ze zvolených metod a pro obě matice byly vždy provedeny tři kroky, které budou v následujícím textu označovány jako experimenty 1, 2 a 3. U každého z experimentů byla použita korelační matice, nikdy kovarianční. Hodnota maximálního počtu iterací pro konvergenci byla ponechána na defaultní hodnotě 25. Software si určí nejvhodnější počet iterací automaticky, v modelu je vždy uvedeno, kolik iterací bylo použito. Jako rotační metoda je používána metoda Varimax. Pro přehlednost jsou na obrázcích výsledných tabulek skryty hodnoty faktorových zátěží nižší než 0,3. Příslušnost původní proměnné k nově vzniklému faktoru je uvažována podle hodnot maximálních faktorových zátěží.

Při prvním experimentu byly faktory extrahovány pomocí vlastních čísel, výstupní faktory měly hodnotu vlastního čísla větší než 1.

U druhého experimentu byl počet latentních proměnných snižován oproti prvnímu pokusu, o čtyři (F-4) a o pět (F-5) faktorů. Bylo zjišťováno, zda půjdou modely vytvořit a jaké budou dávat výsledky.

Třetí experiment zkoumá, jaký vliv má odebrání některých proměnných na předcházející modely. Nejprve byly odfiltrovány proměnné, které v modelu s F-4 faktory měly nízké hodnoty komunality po extrakci faktorů (tabulka communalities, hodnoty nižší než 0,3), tj. faktorová analýza pro ně pracovala nejhůře. S novým počtem proměnných byl vytvořen model, kde výstupní faktory měly hodnotu vlastního čísla větší než 1 a nový model s F-4 faktory. Stejně bylo postupováno u modelu s F-5 faktory. Byly odfiltrovány proměnné, které v modelu s F-5 faktory měly nízké hodnoty komunalit (hodnoty $< 0,3$). S novým počtem proměnných byl vytvořen model, kde výstupní faktory měly hodnotu vlastního čísla větší než 1 a nový model s F-5 faktory.

4.3.1 Metoda hlavních os, matice A

V nastavení uzlu PCA/Factor v programu Clementine 10.1 byla extrakční metodou faktorů zvolena metoda hlavních os (Principal axis factoring) a nastavena metoda rotace. V následujícím textu jsou popsány provedené experimenty. Výsledný stream je znázorněn v příloze 5.

Experiment A1

Na základě vlastního čísla >1 bylo extrahováno celkem 9 faktorů, které vysvětlují celkem 73,264 % variability původních dat. Matice faktorových zátěží vypadá po použití rotace následovně (viz. obrázek 5). Barevně jsou vyznačeny buňky vyjadřující příslušnost původní proměnné k nově vzniklému faktoru, pokud je uvažováno přiřazení proměnných k faktoru na základě maximální faktorové zátěže. Pro rotaci bylo použito 13 iterací.

	Factor								
	1	2	3	4	5	6	7	8	9
P1				-.785				.363	
P2	.685			.476					.304
P3	.385		.524			-.451			
P4				-.303		-.423		-.511	
P5					.641				
P6								.757	
P7			.974						
P8			.980						
P9								.530	
P10	.386								
P11		-.358	-.354				.414		.481
P12							.899		
P13							.846		
P14									.725
P15					.387				
P16						.587			
P17							-.541		.521
P18							.343		.768
P19								-.646	.442
P20	.467							.356	
P21								-.864	
P22	.940								
P23	-.820								
P24	.544	.359					.389		
P25	.740	.456							
P26	.782	.429							
P27							.741		
P28		.301	-.353					-.322	
P29	.373							.409	.596
P30							-.321	.675	
P31	.418	.314							.574
P32		.525							
P33		.329					-.618		-.426
P34		.525							-.352
P35	.585								.392
P36	.713								
P37		.451					-.386		
P38		.440						.449	
P39	.383	.645							
P40	.349	.779							
P41		.648							
P42		.447							

Extraction Method: Principal Axis Factoring.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 13 iterations.

Obrázek 5 – Matice faktorových zátěží, experiment A1 [zdroj: vlastní]

Kaiserova normalizace vede ke snížení hodnoty standardních chyb zatížení pro proměnné s malou komunalitou a zvyšuje korelace mezi šikmými faktory.

Experiment A2

Součástí druhého experimentu bylo vytvoření modelů pro F-4 a F-5 faktory. Po provedení experimentů bylo zřejmé, že modely s pěti i se čtyřmi faktory lze vytvořit. Při pěti výsledných faktorech bylo objasněno 58,055 % variability původních dat, u čtyř faktorů pak 52,751 % variability původních dat. Zároveň se každým snížením počtu extrahovaných faktorů zvyšují

korelace mezi těmito faktory. Vzájemné relace mezi původními proměnnými a získanými faktory jsou zaznamenány na následujících řádcích bez faktorových zátěží, nikoliv již na obrázcích jako v předchozím případě.

Při extrakci 5 faktorů bylo k rotaci využito 15 iterací. Nejvyšší faktorové zátěže po rotaci měly pro faktor 1 proměnné – P2, P10, P14, P20, P22, P23, P24, P25, P26, P31, P32, P35, P36, P39, P40, P42; faktor 2 – P1, P4, P9, P12, P13, P16, P21, P29, P33, P34, P37; faktor 3 – P3, P6, P7, P8, P11, P17, P27; faktor 4 – P5, P15, P18, P19, P30, P38; faktor 5 – P28, P41. Proměnná P6, P9, P14 měly faktorové zátěže nižší než 0,3. Proměnná P23 má stejně blízko ke dvěma faktorům, k faktoru 1 (-.600) a k faktoru 5 (.599). Podobně vypadá situace u proměnné P28, kde je hodnota zátěží téměř shodná u faktoru 3 (-.421) a 5 (.423). Všechny ostatní proměnné jsou snadno přiřaditelné pouze k jednomu z faktorů.

Při extrakci 4 faktorů bylo při rotaci aplikováno 6 iterací. Jednotlivé faktory byly syceny především proměnnými: faktor 1 - P2, P10, P14, P20, P22, P23, P24, P25, P26, P31, P35, P36, P39, P40, P41, P42; faktor 2 – P1, P4, P9, P12, P13, P16, P21, P29, P33, P34, P37; faktor 3 – P3, P6, P7, P8, P11, P17, P27; faktor 4 – P5, P15, P18, P19, P28, P30, P32, P38; Proměnné P6, P9 a P14 měly faktorové zátěže nižší než 0,3.

Experiment A3

Nejprve byly odfiltrovány proměnné P4, P6, P9, P10, P14, P15 a P42. Při „nových“ vstupech, kdy byla vstupní matice snížena o 7 výše uvedených proměnných, bylo extrahováno 8 faktorů, jejichž vlastní číslo je větší než 1. Tyto faktory vysvětlují 76,546 % variability vstupních proměnných. Po provedení rotace s jedenácti iteracemi sytí jednotlivé faktory následující proměnné: faktor 1 – P2, P20, P22, P23, P24, P25, P26, P31, P35, P36; faktor 2 – P3, P7, P8, P27; faktor 3 – P32, P34, P37, P38, P39, P40, P41; faktor 4 – P1, P21, P33; faktor 5 – P5, P18, P19, P30; faktor 6 – P16, P17, P29; faktor 7 – P12, P13; faktor 8 – P11

S uvedenými odfiltrovanými sedmi faktory byl opětovně vytvořen model pro 5 faktorů. Rotace využila třináct iterací. Jak je vidět z následujícího soupisu, proměnné u jednotlivých faktorů se v porovnání s experimentem A2 přeskupily: faktor 1 – P2, P20, P22, P23, P24, P25, P26, P35, P36; faktor 2 – P1, P12, P13, P16, P21, P29, P31, P33, P34; faktor 3 – P3, P7, P8, P11, P17, P27; faktor 4 – P28, P32, P34, P37, P38, P39, P40, P41; faktor 5 – P5, P18, P19, P30. Uvedených pět faktorů objasňuje 65,520 % původní variability dat, což je oproti situaci, kdy do analýzy vstupovalo všech 42 proměnných přibližně o 7,5 % více. Za povšimnutí zde stojí proměnná P39, jejíž faktorové zátěže se u faktorů 1 a 4 téměř shodují (.493 a .494).

Stejný postup pro odfiltrování proměnných byl uplatněn i u modelu se 4 extrahovanými faktory. Zde měly nejnižší hodnoty komunalit proměnné P4, P5, P6, P9, P10, P13, P14, P15, P41, P42. Vstupní matice pro modelování má tedy 32 proměnných. Nejdříve byl opět vytvářen model, kdy jsou extrahovány faktory s hodnotou vlastního čísla větší než 1. Tento model se snažil vytvořit 7 faktorů, ale nebylo možno ho dokončit, protože komunalita některých proměnných překročily hodnotu 1. Z toho důvodu byl vytvořen pouze model pro 4 faktory. Tyto faktory vysvětlují 62,327 % variability vstupních dat, což je asi o 7,5 % více než při původních vstupech. Pro rotaci bylo aplikováno sedm iterací. Vztah vytvořených faktorů k původními proměnnými je po rotaci na základě maximálních zátěží následující: faktor 1 – P2, P20, P22, P23, P24, P25, P26, P31, P35, P36, P39, P40; faktor 2 – P1, P12, P16, P21, P29, P33, P34, P37; faktor 3 – P3, P7, P8, P11, P17, P27; faktor 4 – P18, P19, P28, P30, P32, P38. Po bližším seznámení se s vytvořenými faktory je možno konstatovat, že vazby proměnných na faktory jsou shodné jako v případě čtyř faktorů u experimentu A2. Nedošlo k přeskupení proměnných, změnily se pouze faktorové zátěže.

4.3.2 Metoda hlavních komponent, matice A

Princip práce je obdobný jako u metody hlavních os. V nastavení uzlu byla vybrána extrakční metoda faktorů Principal Components. Stejně jako v předchozím případě byla maticí vstupních dat korelační matice. Maximální počet iterací byl ponechán na 25, u výsledné tabulky je vždy uvedeno, kolik iterací bylo použito. Z důvodu přehlednosti tabulek obsahujících faktorové zátěže, byly skryty hodnoty nižší než 0,3. Výslední stream si lze prohlédnout v příloze 6.

Experiment B1

Postupem využívajícím vlastních čísel bylo získáno 9 faktorů, které vystihují 73,264 % variability vstupních dat. Matice po použití rotace vypadá následovně (viz. obrázek 6). Při rotaci bylo aplikováno 17 iterací. Barevně jsou vyznačeny buňky vyjadřující příslušnost původní proměnné k nově vzniklému faktoru, pokud je uvažováno přiřazení proměnných faktoru podle maximální faktorové zátěže.

	Component								
	1	2	3	4	5	6	7	8	9
P1			-.703				.442		
P2	.742		.462						
P3	.343	.636				-.335			
P4			-.542				-.569		
P5					.739				
P6							.812		
P7		.955							
P8		.953							
P9							.668		
P10	.457								
P11		-.345		-.359		.385		.507	
P12						.883			
P13						.865			
P14								.856	
P15					.381				-.620
P16			.499			.366			.340
P17			-.640	.400					
P18			.357				.789		
P19			-.315	.427			-.550		
P20	.504	.330					.305		
P21						-.853			
P22	.913								
P23	-.800								
P24	.619		.396	.318					
P25	.802			.380					
P26	.850			.335					
P27		.773							
P28		-.428		.359			-.322		
P29	.423		.456		.457	.360			
P30		-.337			.734				
P31	.517		.457			.343			
P32				.575					
P33			-.780	.301					
P34			-.307	.526					
P35	.649							.414	
P36	.779								
P37						.544			-.370
P38						.477		.496	
P39	.451		-.433	.580					
P40	.438			.735					
P41				.738					
P42	.323			.471					.411

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 17 iterations.

Obrázek 6 – Matice faktorových zátěží, experiment B1 [zdroj: vlastní]

Experiment B2

Z devíti faktorů byl snížen jejich počet na 5 a na 4 latentní proměnné. Pět faktorů vystihuje 58,055 % variability vstupů. Při rotaci bylo využito sedm iterací. Vztah manifestních proměnných k faktorům lze na základě maximálních faktorových zátěží charakterizovat následovně: faktor 1 – P2, P10, P14, P20, P22, P23, P24, P25, P26, P31, P35, P36, P40, P41, P42; faktor 2 – P1, P12, P13, P16, P21, P29, P33, P34, P37, P39; faktor 3 – P3, P7, P8, P11, P17, P27; faktor 4 – P5, P15, P18, P19, P28, P30, P32; faktor 5 – P4, P6, P9, P38. Proměnná P14 měla faktorovou zátěž nižší než 0,3.

Druhý model v rámci tohoto pokusu extrahuje čtyři latentní proměnné, které popisují 52,751 % variability vstupů. Rotace využila šesti iterací. Faktory sytí proměnné: faktor 1 – P2, P10, P14, P20, P22, P23, P24, P25, P26, P31, P35, P36, P39, P40, P41, P42; faktor 2 – P1, P4, P9, P12, P13, P16, P21, P29, P33, P34, P37; faktor 3 – P3, P6, P7, P8, P11, P17, P27; faktor 4 – P5, P15, P18, P19, P28, P30, P32, P38. Proměnná P14 měla faktorovou zátěž nižší než 0,3. Pozorované závislosti jsou shodné s metodou hlavních os.

Experiment B3

Na základě modelu s pěti faktory byly odfiltrovány proměnné P14 a P42. Model s faktory generovanými na základě vlastních čísel dá ve výsledku 8 faktorů. Po rotaci využívající 16 iterací vypadají vazby následně: faktor 1 – P2, P10, P20, P22, P24, P25, P26, P31, P35, P36, P39, P40; faktor 2 – P1, P16, P21, P33; faktor 3 – P3, P7, P8, P17, P27, P28; faktor 4 – P5, P15, P18, P19, P29, P30; faktor 5 – P32, P34, P37, P41; faktor 6 – P11, P12, P13; faktor 7 – P4, P6, P9, P38; faktor 8 – P23.

Nově vytvořený model pro pět faktorů se 40 proměnnými vysvětluje 59,998 % variability vstupů. Při rotaci je použito šest iterací. Nejvyšší faktorové zátěže měly: faktor 1 – P2, P10, P20, P22, P23, P24, P25, P26, P31, P35, P36, P40, P41; faktor 2 – P1, P12, P13, P16, P21, P29, P33, P34, P37, P39; faktor 3 – P3, P7, P8, P11, P17, P27; faktor 4 – P5, P15, P18, P19, P28, P30, P32; faktor 5 – P4, P6, P9, P38. Výsledné vztahy jsou stejné jako u experimentu B2, jen s tím rozdílem, že zde nejsou zahrnuty odfiltrované proměnné.

Na základě modelu se čtyřmi faktory byly odfiltrovány proměnné P6, P9, P14, P15 a P42. Model s faktory generovanými na základě vlastních čísel dá ve výsledku 8 faktorů. Po rotaci využívající 15 iterací vypadají vazby následně: faktor 1 – P2, P10, P20, P22, P24, P25, P26, P31, P35, P36, P40; faktor 2 – P3, P7, P8, P17, P27; faktor 3 – P4, P16, P21, P29, P33, P39; faktor 4 – P32, P34, P37, P38, P41; faktor 5 – P5, P18, P19, P30; faktor 6 – P12, P13; faktor 7 – P23, P28; faktor 8 – P1, P11.

Nově vytvořený model pro čtyři faktory s 37 proměnnými vysvětluje 57,751 % variability vstupů. Při rotaci je použito sedm iterací. Faktory sytí především následující proměnné: faktor 1 – P2, P10, P20, P22, P23, P24, P25, P26, P31, P35, P36, P40; faktor 2 – P1, P4, P12, P13, P16, P21, P29, P33, P34, P37, P39; faktor 3 – P3, P7, P8, P11, P17, P27; faktor 4 – P5, P18, P19, P28, P30, P32, P38, P41. Oproti modelu v druhém pokusu došlo k několika změnám.

4.3.3 Metoda hlavních os, matice B

Je zde použit stejný postup práce jako u předchozích metod pro matici A. Konečný vzhled streamu si lze prohlédnout v příloze 7.

Experiment C1

Na základě sestaveného modelu bylo vygenerováno celkem 9 faktorů, které dohromady vysvětlují 73,272 % variability původních dat. Matice po použití rotace vypadá následovně (viz. obrázek 7). Pro rotaci bylo použito 14 iterací.

	Factor								
	1	2	3	4	5	6	7	8	9
P1						-.806		.343	
P2	.676					.481			.327
P3	.369		.516	-.441					
P4				-.454				-.509	
P5					.610				
P6								.776	
P7			.963						
P8			.974						
P9								.537	
P10	.362								
P11		-.377	-.315				.393		.491
P12							.910		
P13							.829		
P14									.724
P15					.403				
P16				.618					
P17								-.526	.543
P18	-.319							.341	.747
P19								.449	-.675
P20	.423							.377	.350
P21									-.854
P22	.940								
P23	-.834								
P24	.538	.354							.380
P25	.729	.456							
P26	.767	.445							
P27								.642	
P28		.312							-.360
P29	.377							.605	.412
P30									.715
P31	.422	.318						.581	
P32		.531							
P33		.337						-.498	-.557
P34		.512						-.328	
P35		.577							.401
P36		.707							
P37							.437	-.374	
P38							.445		.444
P39		.374	.651						
P40		.336	.784						
P41			.642						
P42			.448						

Extraction Method: Principal Axis Factoring
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 14 iterations.

Obrázek 7 – Matice faktorových zátěží, experiment C1 [zdroj: vlastní]

Experiment C2

Z devíti faktorů byl jejich počet snížen na 5 a na 4 faktory. Pět faktorů vystihuje 58,047 % variability vstupů. Při rotaci bylo využito šest iterací. Vztah manifestních proměnných k faktorům lze na základě maximálních faktorových zátěží vystihnout následovně: faktor 1 – P2, P10, P14, P20, P22, P23, P24, P25, P26, P31, P35, P36, P40, P41, P42; faktor 2 – P1, P12, P13, P16, P21, P29, P33, P34, P37, P39; faktor 3 – P3, P7, P8, P11, P17, P27; faktor 4 – P5, P15, P18, P19, P28, P30, P32; faktor 5 – P4, P6, P9, P38. Proměnná P14 měla faktorovou zátěž nižší než 0,3. Výstup je shodný s metodou hlavních komponent pro matici A.

Druhý model v rámci tohoto pokusu extrahuje čtyři latentní proměnné, které popisují 52,597 % variability vstupů. Rotace využila šesti iterací. Spojitost faktorů s původními proměnnými je asi následující: faktor 1 – P2, P10, P14, P20, P22, P23, P24, P25, P26, P31, P35, P36, P39, P40, P41, P42; faktor 2 – P1, P4, P9, P12, P13, P16, P21, P29, P33, P34, P37; faktor 3 – P3, P6, P7, P8, P11, P17, P27; faktor 4 – P5, P15, P18, P19, P28, P30, P32, P38. Proměnné P6, P9, P14 měla faktorovou zátěž nižší než 0,3. Výstup je shodný s metodou hl. os i hlavních komponent pro matici A.

Experiment C3

Na základě modelu s pěti faktory byly odfiltrovány proměnné P5, P10, P14, P15, P41 a P42. Z modelu s faktory generovanými na základě vlastních čísel vystupuje celkem 8 faktorů. Po rotaci aplikující 10 iterací vypadají vazby následně: faktor 1 – P2, P24, P25, P26, P31, P35, P36, P39, P40; faktor 2 – P1, P4, P16, P21, P29, P33, P34, P37; faktor 3 – P3, P7, P8, P17, P20, P27; faktor 4 – P18, P19, P28, P30; faktor 5 – P12, P13; faktor 6 – P6, P9, P38; faktor 7 – P22, P23; faktor 8 – P11, P32.

Nově vytvořený model pro pět faktorů se 36 proměnnými vysvětluje 63,603 % variability vstupů. Při rotaci je použito sedm iterací. Spojitost faktorů s manifestními proměnnými je podle maximálních faktorových zátěží takovýto: faktor 1 – P2, P20, P22, P23, P24, P25, P26, P31, P35, P36, P40; faktor 2 – P1, P12, P13, P16, P21, P29, P33, P34, P37, P39; faktor 3 – P3, P7, P8, P11, P17, P27; faktor 4 – P18, P19, P28, P30, P32; faktor 5 – P4, P6, P9, P38. Výsledné vztahy jsou stejné jako u experimentu C2, jen s tím rozdílem, že zde nejsou zahrnuty odfiltrované proměnné.

Souhlasně s modelem se čtyřmi faktory v experimentu C2 byly odfiltrovány proměnné P4, P5, P6, P9, P10, P13, P14, P15, P41 a P42. Model s faktory generovanými na základě vlastních čísel dá ve výsledku 7 faktorů, které vysvětlují 74,917 % variability vstupních proměnných. Po rotaci využívající 24 iterací vypadají vazby následně: faktor 1 – P2, P20,

P22, P24, P25, P26, P31, P32, P35, P36, P39, P40; faktor 2 – P11, P12, P16, P17, P29, P33, P34, P37; faktor 3 – P3, P7, P8, P27, P33, P39; faktor 4 – P1, P21; faktor 5 – P18, P19, P30; faktor 6 – P28, P38; faktor 7 – P23.

Nově vytvořený model pro čtyři faktory s 32 proměnnými vysvětluje 62,212 % variability vstupů. Při rotaci je použito šest iterací. Faktory sytí především proměnné: faktor 1 – P2, P20, P22, P23, P24, P25, P26, P31, P35, P36, P39, P40; faktor 2 – P1, P12, P16, P21, P29, P33, P34, P37; faktor 3 – P3, P7, P8, P11, P17, P27; faktor 4 – P18, P19, P28, P30, P32, P38. Oproti modelu v druhém pokusu došlo k několika změnám.

4.3.4 Metoda hlavních komponent, matice B

Stejně jako v předchozích případech i zde byly vypracovány tři experimenty. Uzly použité pro jejich zpracování jsou zobrazeny na streamu v příloze 8.

Experiment D1

Postupem využívajícím vlastních čísel bylo získáno 9 faktorů, které vystihují 73,272 % variability vstupních dat. Matice po použití rotace vypadá následovně (viz. obrázek 8). Při rotaci bylo aplikováno 16 iterací. Barevně jsou vyznačeny buňky vyjadřující příslušnost původní proměnné k nově vzniklému faktoru, pokud je uvažováno přiřazení proměnných faktorů podle maximální faktorové zátěže.

	Component								
	1	2	3	4	5	6	7	8	9
P1		-.702					.449		
P2	.711	.479							
P3	.358			.634		-.320			
P4		-.546					-.563		
P5					.704				
P6							.805		
P7				.948					
P8				.944					
P9							.685		
P10	.395								
P11			-.374	-.314		.410		.484	
P12						.873			
P13						.859			
P14								.863	
P15									.769
P16		.546				.407			
P17		.426				-.633			
P18						.364	.750		
P19		.445					-.560	.317	
P20	.416					.449			.404
P21		-.847							
P22	.917								
P23	-.824								
P24	.603	.411	.340						
P25	.763		.402						
P26	.805		.379						
P27			.304	.668					
P28			.352	-.319	-.316	-.350			
P29	.410	.476			.444	.351			
P30					.800				
P31	.483	.499	.309				.347		
P32			.594						
P33		-.778	.302						
P34		-.306	.524						
P35	.616							.449	
P36	.783								
P37						.482			
P38						.475		.480	
P39	.422	-.414	.606						
P40	.386		.745						
P41			.731						
P42	.319		.526						

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 16 iterations.

Obrázek 8 – Matice faktorových zátěží, experiment D1 [zdroj: vlastní]

Experiment D2

Z devíti faktorů extrahovaných v experimentu D1 byl počet snížen stejně jako v předcházejících experimentech A2, B2 i C2 na 5 a na 4 faktory. Pět faktorů vystihuje 58,047 % variability vstupů. Při rotaci bylo využito sedm iterací. Vztah manifestních proměnných k faktorům lze na základě maximálních faktorových zátěží určit následovně: faktor 1 – P2, P10, P20, P22, P23, P24, P25, P26, P31, P35, P36, P39, P40, P42; faktor 2 – P1, P12, P13, P16, P21, P29, P33, P34, P37; faktor 3 – P3, P7, P8, P11, P17, P27, P41; faktor 4 – P5, P15, P18, P19, P30, P32; faktor 5 – P4, P6, P9, P14, P28, P38.

Druhý model v rámci tohoto experimentu extrahuje čtyři latentní proměnné, které popisují 52,597 % variability vstupů. Rotace využila šesti iterací. Relace mezi faktory a proměnnými jsou: faktor 1 – P2, P10, P14, P20, P22, P23, P24, P25, P26, P31, P35, P36, P39, P40, P41, P42; faktor 2 – P1, P4, P9, P12, P13, P16, P21, P29, P33, P34, P37; faktor 3 – P3, P6, P7, P8, P11, P17, P27; faktor 4 – P5, P15, P18, P19, P28, P30, P32, P38. Získané faktory jsou naprosto shodné s modelem B2, liší se pouze hodnotami faktorových zátěží.

Experiment D3

S ohledem na model s pěti faktory v experimentu D2 byly odfiltrovány proměnné P14 a P42. Model generující faktory na základě vlastních čísel dá 9 latentních proměnných. Po rotaci využívající 10 iterací vypadají vazby následovně: faktor 1 – P2, P10, P22, P23, P24, P25, P26, P31, P35, P36, P39, P40; faktor 2 – P1, P16, P21, P29, P33; faktor 3 – P3, P7, P8, P17, P20, P27; faktor 4 – P32, P34, P37, P41; faktor 5 – P5, P18, P19, P30; faktor 6 – P11, P12, P13; faktor 7 – P4, P6, P9, P38; faktor 8 – P28; faktor 9 – P15.

Nově vytvořený model pro pět faktorů se 40 proměnnými vysvětluje 59,967 % variability vstupů. Při rotaci je použito sedm iterací. Faktory sytí především proměnné: faktor 1 – P2, P10, P20, P22, P23, P24, P25, P26, P31, P35, P36, P40; faktor 2 – P1, P4, P12, P13, P16, P21, P29, P33, P34, P37, P39; faktor 3 – P3, P7, P8, P11, P17, P27, P41; faktor 4 – P5, P15, P18, P19, P30, P32; faktor 5 – P6, P9, P28, P38.

Na základě modelu se čtyřmi faktory byly odfiltrovány proměnné P4, P5, P6, P9, P10, P14, P15 a P42. Model s faktory generovanými na základě vlastních čísel dá 8 faktorů, které objasňují 77,182 % variability vstupních dat. Po rotaci využívající 13 iterací vypadají nejsilnější vazby následně: faktor 1 – P2, P24, P25, P26, P31, P35, P36, P40; faktor 2 – P3, P7, P8, P17, P20, P27; faktor 3 – P1, P16, P21, P29, P33, P39; faktor 4 – P18, P19, P28, P30; faktor 5 – P32, P34, P38, P41; faktor 6 – P12, P13, P19; faktor 7 – P23, P28; faktor 8 – P11,

Nově vytvořený model pro čtyři faktory s 34 proměnnými vysvětluje 60,335 % variability vstupů. Při rotaci je použito šest iterací. Relace faktorů s proměnnými vypadá následovně: faktor 1 – P2, P22, P23, P24, P25, P26, P31, P35, P36, P39; faktor 2 – P1, P12, P13, P16, P19, P21, P29, P33, P34; faktor 3 – P3, P7, P8, P11, P17, P18, P20, P27; faktor 4 – P28, P30, P32, P37, P38, P40, P41.

4.4 Interpretace výstupů

4.4.1 Experiment 1

Při analýze experimentu 1 byly brány v úvahu všechny modely, tedy A1, B1, C1 a D1. Poměrně složitým statisticko-matematickým výpočtem, při jehož zpracování byl použit program Clementine verze 10.1 bylo extrahováno devět nejdůležitějších faktorů, jež společně vysvětlují 73,264 % (model A1 a B1) – 73,272 % (model C1 a D1) celkové variability souboru proměnných. Současně byly identifikovány proměnné použitelné při interpretaci jednotlivých faktorů (vyznačeny na obrázcích 5 až 8 šedě). Pro interpretaci výstupů je použit model C1, protože vysvětluje nepatrně více celkové variability souboru proměnných, než modely A1 a B1, současně se od modelu A1 liší pouze přiřazením 3 proměnných (P17, P28 a P38) k jiným faktorům. Je třeba dodat, že modely se vzájemně výrazně nelišily. Rozdíly byly zejména v číslování faktorů, např. proměnné náležející v modelu A1 k faktoru 2 byly v modelu B1 přiřazeny faktoru 4 nebo proměnné náležející v modelu A1 k faktoru 8 byly v modelu B1 a D1 spojeny s faktorem 7 apod. Rotace u modelu C1 usnadnila interpretaci a snížila procento objasňované variability o 6,597 % na 66,675 %.

První faktor vysvětluje po rotaci 13,976 % celkové variability souboru proměnných. Je charakteristický vysokými zátěžemi následujících proměnných: P22-Lékaři na 1000 obyvatel, P23-Počet obyvatel na 1 lékaře, P26-Průměrný počet nemocensky pojištěných, P25-Lékárny a výdejny léků, P36-Divadla, P2-Hustota obyvatelstva, P35-Galerie, P24-Nemocnice, P20-Volná pracovní místa a P10-Pořízené investice na ochranu životního prostředí. Pro tento faktor by mohlo být použito označení Faktor zdraví.

Rotovaný druhý faktor vysvětluje 9,732 % celkové variability souboru proměnných. Vysokých zátěží je dosahováno u následujících proměnných: P40-Tělocvičny, P39-Hřiště, P41-Stadiony včetně krytých, P32-Stálá kina, P34-Muzea, P42-Zimní stadiony včetně krytých, P38-Koupaliště a bazény, P37-Přírodní amfiteátry. Název vystihující „obsah“ faktoru může být Faktor relaxace.

Třetí faktor objasňuje 8,545 % celkové variability souboru proměnných. Vysokých zátěží je dosahováno u následujících proměnných: P8-Průměrný měsíční důchod starobní, P7-Průměrný měsíční důchod celkem, P27-Průměrná doba trvání 1 případu pracovní neschopnosti, P3-Průměrný věk obyvatel. Tento faktor vystihuje pojmenování Sociální situace.

Čtvrtý extrahovaný faktor vysvětlující 7,198 % celkové variability souboru proměnných má vysoké faktorové zátěže u proměnných: P16-Rozvody, P29-Zjištěné trestné

činy, P31-Požáry, P17-Potravy. Název, který by shrnoval obsažené proměnné, může být vyjádřen termínem Duševní pohoda.

Pátý faktor objasňuje 6,890 % celkové variability souboru proměnných. Vysoké faktorové zátěže zde dosahují proměnné P18-Celkový přírůstek obyvatelstva, P30-Dopravní nehody, P19-Míra registrované nezaměstnanosti, P5-Dokončené byty, P15-Sňatky, P28-Průměrné procento pracovní neschopnosti za rok. Za nejuvýstižnější pojmenování tohoto faktoru je považována Koncentrace obyvatel.

Šestý rotovaný faktor vysvětluje 6,575 % celkové variability souboru proměnných. Vysokých zátěží je dosahováno u proměnných: P21-Délka silnic a dálnic, P1-Rozloha okresu, P33-Veřejné knihovny včetně poboček. Název faktoru, který vystihuje zahrnuté proměnné, je Rozloha území.

Sedmý extrahovaný faktor vysvětluje 5,208 % celkové variability souboru proměnných. Vysoké zátěže jsou charakteristické pro proměnné: P12-Emise základních znečišťujících látek do ovzduší - oxid siřičitý, P13-Emise základních znečišťujících látek do ovzduší - oxidy dusíku. Vzhledem k tomu, že obě proměnné souvisejí s emisí znečišťujících látek do ovzduší, je pojmenován Znečištěné ovzduší.

Poslední dva faktory vysvětlují méně než 5 % celkové variability souboru proměnných, v pořadí osmý faktor objasňuje 4,996 % variability a devátý faktor 3,555 % variability proměnných. Vysoké zátěže u osmého faktoru jsou typický pro proměnné P6-Hromadná ubytovací zařízení, P9-Maloplošná chráněná území, P4-Podíl vybraných druhů pozemků - zemědělská půda, zatímco u devátého faktoru se jedná o proměnné P14-Emise základních znečišťujících látek do ovzduší - oxid uhelnatý a P11-Emise základních znečišťujících látek do ovzduší - emise tuhé. V pořadí osmý faktor může nést název Atraktivita území pro turisty, zatímco poslední faktor Znečištění ovzduší v důsledku procesu spalování.

4.4.2 Experiment 2

Ve všech čtyřech případech byly v experimentu 2 vytvořeny pět a čtyři faktory. Objasněný podíl variability souboru proměnných se po rotaci liší podle použité metody extrakce faktorů. V případě, že byla použita metoda hlavních os (modely A1 a C1), pak extrahované faktory vysvětlují po rotaci přibližně 53 % variability souboru proměnných v situaci, kdy bylo generováno pět latentních proměnných. Pro čtyři latentní proměnné se procento vysvětlení variability pohybuje lehce nad hranicí 48 % celkové variability souboru.

Lepší výsledky představovalo použití metody hlavních komponent, kde pět faktorů vysvětlovalo po aplikaci rotace Varimax zhruba 58 % a čtyři faktory okolo 52 % variability souboru proměnných. Pro interpretaci se tedy jeví jako lepší varianta využití modelu B2 nebo D2. Byl zvolen model B2, protože obsahuje data za 9 let, naproti tomu model D2 pouze za 8 let. Z výstupu, který software poskytuje, je patrné, že se zároveň s každým snížením počtu extrahovaných faktorů zvyšují korelace mezi těmito faktory. Pořadí proměnných v rámci faktorů je seřazeno sestupně podle výše faktorových zátěží v absolutní hodnotě.

V rámci experimentu je nejdříve popsán model, kdy bylo extrahováno pět faktorů. V tomto případě první faktor vysvětluje po rotaci 19,165 % celkové variability souboru proměnných. Charakteristické jsou pro něho vysoké zátěže následujících proměnných: P26-Průměrný počet nemocensky pojištěných, P25-Lékárny a výdejny léků, P2-Hustota obyvatelstva, P22- Lékaři na 1000 obyvatel, P36-Divadla, P24-Nemocnice, P35-Galerie, P40-Tělocvičny, P31-Požáry, P23-Počet obyvatel na 1 lékaře, P20-Volná pracovní místa, P10-Pořízené investice na ochranu životního prostředí, P42-Zimní stadiony včetně krytých, P41-Stadiony včetně krytých a P14-Emise základních znečišťujících látek do ovzduší - oxid uhelnatý s faktorovou zátěží nižší než 0,3. Vhodným pojmenováním se jeví Zdravotně-volnočasový faktor.

Druhý faktor vysvětluje 13,525 % celkové variability souboru proměnných. Vysokých zátěží je dosahováno u následujících proměnných: P1-Rozloha okresu, P12-Emise základních znečišťujících látek do ovzduší - oxid siřičitý, P13-Emise základních znečišťujících látek do ovzduší - oxidy dusíku, P16-Rozvody, P21-Délka silnic a dálnic, P29-Zjištěné trestné činy, P33-Veřejné knihovny včetně poboček, P34-Muzea, P37-Přírodní amfiteátry a P39-Hřiště. Jednoslovný název pro pojmenování faktoru je obtížné určit, proto je použito víceslovné, ale výstižnější: Faktor životního prostředí a kultury.

Třetí faktor objasňuje 10,302 % celkové variability souboru proměnných. Vysokých zátěží je dosahováno u následujících proměnných: P3-Průměrný věk obyvatel, P7-Průměrný měsíční důchod celkem, P8-Průměrný měsíční důchod starobní, P11-Emise základních znečišťujících látek do ovzduší - emise tuhé, P17-Potravy, P27-Průměrná doba trvání 1 případu pracovní neschopnosti. Za termín vystihující tento faktor je považován název Sociální stav.

Čtvrtý extrahovaný faktor vysvětluje 8,902 % celkové variability souboru proměnných má vysoké faktorové zátěže u proměnných: P5-Dokončené byty, P15-Sňatky, P18-Celkový přírůstek obyvatelstva, P19-Míra registrované nezaměstnanosti, P28-Průměrné

procento pracovní neschopnosti za rok, P30-Dopravní nehody, P32-Stálá kina. Název, kterým by mohl být faktor označen, je Koncentrace obyvatel.

Pátý a poslední faktor objasňuje 6,161 % celkové variability souboru proměnných. Vysoké faktorové zátěže zde dosahují proměnné: P4-Podíl vybraných druhů pozemků - zemědělská půda, P6-Hromadná ubytovací zařízení a P9-Maloplošná chráněná území, P38-Koupaliště a bazény. Pojmenování tohoto faktoru může být Turistická atraktivita.

V druhé fázi byly generovány čtyři faktory. Pro tuto situaci platí, že první faktor vysvětluje 18,963 % celkové variability souboru proměnných a vysokých faktorových zátěží je dosahováno u proměnných P26-Průměrný počet nemocensky pojištěných, P25-Lékárny a výdejny léků, P2-Hustota obyvatelstva, P22- Lékaři na 1000 obyvatel, P36-Divadla, P24-Nemocnice, P35-Galerie, P40-Tělocvičny, P31-Požáry, P23-Počet obyvatel na 1 lékaře, P39-Hřiště, P20-Volná pracovní místa, P10-Pořízené investice na ochranu životního prostředí, P42-Zimní stadiony včetně krytých, P41-Stadiony včetně krytých a P14-Emise základních znečišťujících látek do ovzduší - oxid uhelnatý s faktorovou zátěží nižší než 0,3. Název tohoto faktoru může být Zdravotně-volnočasový faktor, stejně jako v předcházejícím modelu, kdy bylo extrahováno pět faktorů.

Druhý rotovaný faktor objasňuje 13,672 % variability souboru proměnných. Faktor sytí především proměnné: P33-Veřejné knihovny včetně poboček, P21-Lékaři na 1000 obyvatel, P12-Emise základních znečišťujících látek do ovzduší - oxid siřičitý, P16-Rozvody, P1-Rozloha okresu, P29-Zjištěné trestné činy, P13-Emise základních znečišťujících látek do ovzduší - oxidy dusíku, P34-Muzea, P4-Podíl vybraných druhů pozemků - zemědělská půda, P37-Přírodní amfiteátry a P9-Maloplošná chráněná území. Také u tohoto faktoru může být použito označení aplikované v předešlém modelu a to Faktor životního prostředí a kultury.

Třetí faktor objasňuje po rotaci 10,825 % variability souboru proměnných. Faktor sytí především proměnné: P8-Průměrný měsíční důchod starobní, P7-Průměrný měsíční důchod celkem, P27-Průměrná doba trvání 1 případu pracovní neschopnosti, P3-Průměrný věk obyvatel, P17-Potravy, P11-Emise základních znečišťujících látek do ovzduší - emise tuhé, P6-Hromadná ubytovací zařízení. Za pojmenování, které nejvíce vystihuje obsah faktoru je považován název Sociální situace.

Čtvrtý faktor vysvětlující 9,291 % celkové variability souboru proměnných má vysoké faktorové zátěže u proměnných: P30-Dopravní nehody, P19-Míra registrované nezaměstnanosti, P18-Celkový přírůstek obyvatelstva, P5-Dokončené byty, P28-Průměrné procento pracovní neschopnosti za rok, P15-Sňatky, P38-Koupaliště a bazény a P32-Stálá

kina. Za adekvátní označení tohoto faktoru je považováno Koncentrace obyvatel, název, který byl již dříve také použit.

Interpretace faktorů v druhém experimentu byla daleko náročnější a komplikovanější než u prvního experimentu, stejně tak vymýšlení výstižných názvů faktorů bylo obtížnější pro model se čtyřmi faktory. Důvod je ten, že zdejší faktory jsou vytvořeny kombinací faktorů, které byly v předchozím experimentu či modelu s pěti faktory samostatné nebo lépe uspořádané. Z uvedeného vyplývá, že pokud není počet extrahovaných faktorů striktně určen, je z hlediska interpretace vhodnější vytvoření více faktorů, jejichž počet je možné na základě experimentů snižovat do té doby, dokud jsou faktory bez větších obtíží popsateľné, než bezdůvodně vytvářet model s malým množstvím faktorů, které jsou velmi obtížně komentovatelné.

4.4.3 Experiment 3

V rámci vyhodnocení experimentu 3 již nejsou specifikovány jednotlivé faktory a proměnné, které je sytí, ale bylo provedeno porovnání toho, které proměnné byly odfiltrovány v závislosti na použité metodě a rozměru vstupní matice. Na základě těchto údajů bylo provedeno shrnutí. Jak již bylo řečeno dříve, proměnné byly odfiltrovány na základě nízké hodnoty komunalit. Za nízkou hodnotou byla považována hodnota komunality po extrakci faktorů nižší než 0,3. Podle této specifikace byly v modelech pro pět a pro čtyři faktory z experimentu 2 odfiltrovány proměnné znázorněné v tabulce 2 a tvořeny nové modely, které jsou popsány v rámci každého experimentu.

Tabulka 2 – Odfiltrované proměnné [zdroj: vlastní]

Model	Odfiltrované proměnné pro 5 extrahovaných faktorů	Odfiltrované proměnné pro 4 extrahované faktory
A3	P4, P6, P9, P10, P14, P15, P42	P4, P5, P6, P9, P10, P13, P14, P15, P41, P42
B3	P14, P42	P6, P9, P14, P15, P42
C3	P5, P10, P14, P15, P41, P42	P4, P5, P6, P9, P10, P13, P14, P15, P41, P42
D3	P14, P42	P4, P5, P6, P9, P10, P14, P15, P42

Z pohledu do tabulky 2 je patrné, že počet filtrovaných proměnných byl u modelů, kdy bylo extrahováno pět faktorů, nižší, než v případě extrakce čtyř faktorů. Stejně tak bylo

odfiltrováno méně proměnných, pokud byla při extrakci faktorů použita metoda hlavních komponent, což bylo v případě modelů B3 a D3. Bude-li věnována pozornost nejprve modelu pro pět faktorů, byly pokaždé vyřazeny proměnné P14-Emise základních znečišťujících látek do ovzduší - oxid uhelnatý a P42-Zimní stadiony včetně krytých, dvakrát ze čtyř případů byly odfiltrovány proměnné P10-Pořízené investice na ochranu životního prostředí, P14-Emise základních znečišťujících látek do ovzduší - oxid uhelnatý a P15-Sňatky, pouze v jednom ze čtyř modelů byly vyloučeny proměnné P4-Podíl vybraných druhů pozemků - zemědělská půda, P5-Dokončené byty, P6-Hromadná ubytovací zařízení, P9-Maloplošná chráněná území a P41-Stadiony včetně krytých. Naproti tomu v modelech se čtyřmi extrahovanými faktory bylo ve všech čtyřech případech odfiltrováno pět proměnných, opět to byly proměnné P14-Emise základních znečišťujících látek do ovzduší - oxid uhelnatý a P42-Zimní stadiony včetně krytých, ale přibýly k nim ještě P6-Hromadná ubytovací zařízení, P9-Maloplošná chráněná území a P15-Sňatky. Další skupinu ukazatelů, které byly vyřazeny ve třech ze čtyř možných případů, tvoří P4-Podíl vybraných druhů pozemků - zemědělská půda, P5-Dokončené byty a P10-Pořízené investice na ochranu životního prostředí. Ve dvou případech, pokaždé při použití metody hlavních os pro extrakci faktorů, byly odstraněny proměnné P13-Emise základních znečišťujících látek do ovzduší - oxidy dusíku a P14-Emise základních znečišťujících látek do ovzduší - oxid uhelnatý.

Z tabulky 2 je možno též vyzorovat, že nejméně proměnných s komunalitou nižší než 0,3, tj. odfiltrovaných proměnných, vstupovalo do modelů v experimentu B3. Za ním se na další pomyslné příčce umístily modely v experimentu D3. V obou případech byla pro extrakci faktorů použita metoda hlavních komponent, společnými faktory jí je tedy vysvětlována větší část variability či proměnlivosti proměnných.

5 Závěr

Začátek diplomové práce byl věnován seznámení s oblastí kvality života, zmíněno bylo její hodnocení na základě ukazatele HDI, který je využíván OSN k hodnocení kvality života členských států, a faktové analýzy. V další části se již diplomová práce věnovala modelování kvality života pomocí faktorové analýzy. Bylo ověřeno, že faktorovou analýzu pro zpracování dané problematiky použít lze.

Byl definován problém P, jehož složky tvoří kvalita života a faktorová analýza a byl stanoven postup vedoucí k naplnění cíle. Sestavila jsem datovou matici a vybrala atributy vstupující do modelu. Matice vstupních dat byly nakonec vytvořeny dvě. Matice A pro roky 2001-2009, kde byly hodnoty jedenácti proměnných v roce 2009 doplněny pomocí funkce medián a matice B pro roky 2001-2008. K datové matici jsem vytvořila datový slovník, který obsahuje základní charakteristiku proměnných.

Modelování proběhlo v programu Clementine verze 10.1. Pro extrakci faktorů byly použity dvě metody, metoda hlavních os a metoda hlavních komponent. Pro modelování byly použity matice A i B. Kombinací metody získávání faktorů a matice byly vytvořeny čtyři modely. V rámci těchto modelů jsem provedla vždy tři experimenty. Výstupy byly analyzovány a interpretovány.

Počet faktorů byl v experimentu 1 určen pomocí vlastních čísel, extrahovány byly faktory s vlastním číslem větším než jedna. Tento postup vygeneroval v každém experimentu, tj. A1, B1, C1 a D1 vždy 9 faktorů, které se lišily pouze minimálně. Pro interpretaci výstupů byl použit model C1. Faktory byly v rámci interpretace výstupů pojmenovány. Snahou bylo, aby názvy odrážely v co největší míře proměnné, které faktory sytí. Faktory byly nazvány: Faktor zdraví, Faktor relaxace, Sociální situace, Duševní pohoda, Koncentrace obyvatel, Rozloha území, Znečištěné ovzduší, Atraktivita území pro turisty a Znečištění ovzduší v důsledku procesu spalování.

V rámci druhých experimentů, tj. A2, B2, C2 a D2 byl počet faktorů snížen o pět a o čtyři oproti počtu faktorů vygenerovaných metodou vlastních čísel a bylo pozorováno, k jakým změnám dochází. Ve všech čtyřech případech byly v experimentu 2 vytvořeny pět a čtyři faktory. I v tomto případě byly faktory pojmenovány, názvy jsou: Zdravotně-volnočasový faktor, Faktor životního prostředí a kultury, Sociální stav, Koncentrace obyvatel a Turistická atraktivita. Označení čtyř výstupních faktorů je téměř stejné.

Interpretace faktorů získaných v experimentu 2 byla mnohem obtížnější a komplikovanější než u experimentu předchozího. Stejně tak je složitější popsat 4 faktory než

5 faktorů, názvy faktorů již v takové míře nevystihují obsažené proměnné. Z uvedeného vyplývá, že není-li počet extrahovaných faktorů pevně stanoven, je z hlediska interpretace vhodnější vytvoření více faktorů, jejichž počet je možné na základě experimentů snižovat do doby, dokud jsou faktory bez větších obtíží popsatelné, než vytvářet model s malým množstvím faktorů, které jsou velmi obtížně komentovatelné.

Ve třetích experimentech byly vytvářeny modely, kde byly některé proměnné odfiltrovány. Byly vyřazeny ty proměnné, které měly po extrakci faktorů nízké hodnoty komunalit, tj. část variability proměnných vysvětlená společnými faktory byla nižší než 0,3. Je žádoucí, aby byl rozdíl komunalit na počátku a po extrakci pouze zvolených faktorů co nejmenší. Ve výsledku bylo porovnáno, které proměnné měly nejčastěji nízké hodnoty komunalit a u kterých metod. Bylo zjištěno, že méně proměnných s nízkou hodnotou komunalit dává v modelech s pěti i se čtyřmi faktory metoda hlavních komponent.

Seznam zkratk

HDI	Index lidského rozvoje
OSN	Organizace spojených národů
PCA	Metoda hlavních komponent
KŽ	Kvalita života

6 Použité zdroje

- [1] AKRANA VIČIŮTĚ, D.; RUŽEVIČIUS, J. Quality of Life and its Components' Measurement. *ENGINEERING ECONOMICS* [online]. 2007, No 2, [cit. 2009-10-23]. Dostupný z WWW: <<http://web.ebscohost.com/ehost/pdf?vid=4&hid=2&sid=199fd247-f38c-4e61-9936-3aff1c300690%40sessionmgr3>>. ISSN 1392-2785.
- [2] BLAHUŠ, P. *Faktorová analýza a její zobecnění*. Praha : SNTL, 1985. 356 s.
- [3] Clementine 10.1 – Help
- [4] *Faktorová analýza* [online]. [s.l.], 2004. 4 s. Doprovodná pomůcka ke studiu. ZCU. Dostupné z WWW: <<http://home.zcu.cz/~alexk/sad1/vyuka/FA-teorie.pdf>>.
- [5] GARSON, G. D. *Factor Analysis : Statnotes, from North Carolina State University, Public Administration Program* [online]. c2010, 11.2.2010 [cit. 2010-03-02]. Factor Analysis. Dostupné z WWW: <<http://faculty.chass.ncsu.edu/garson/PA765/factor.htm#assume>>.
- [6] HEBÁK, P., et al. *Vícerozměrné statistické metody [3]*. druhé, doplněné vydání. Praha : INFORMATORIUM, 2007. 271 s. ISBN 978-80-7333-001-9.
- [7] KŘIVOHLAVÝ, J. Kvalita života. In *Kvalita života : Sborník příspěvků z konference, konané dne 25.10.2004 v Třeboni*. Kostelec nad Černými lesy : INSTITUT ZDRAVOTNÍ POLITIKY A EKONOMIKY, 2004 [cit. 2009-12-18]. Dostupné z WWW: <<https://www.ipvz.cz/download.aspx?item=1105&>>. ISBN 80-86625-20-6.
- [8] KUBANOVÁ, J. *Statistické metody pro ekonomickou a technickou praxi*. Vydání druhé. Bratislava : STATIS, 2004. 249 s. ISBN 80-85659-37-9.
- [9] List of countries by Human Development Index In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida) : Wikipedia Foundation, , 17.4.2010 [cit. 2010-04-20]. Dostupné z WWW: <http://en.wikipedia.org/wiki/List_of_countries_by_Human_Development_Index>.
- [10] MEDERLY, P.; TOPERCER, J.; NOVÁČEK, P. *Indikátory kvality života a udržitelného rozvoje : kvantitativní, vícerozměrný a variantní přístup*. Praha : UK FSV CESES, 2004. 117 s. ISBN 80-239-4389-8.
- [11] MELOUN, M.; MILITKÝ, J. *Kompendium statistického zpracování dat : metody a řešené úlohy včetně CD*. Vydání 1. Praha : Academia, 2002. 764 s. ISBN 80-200-1008-4.
- [12] *OSN Praha* [online]. c2005 [cit. 2010-02-27]. Dostupné z WWW: <<http://www.osn.cz/zpravodajstvi/zpravy/zprava.php?id=1199>>.

- [13] SPSS [online]. 2010 [cit. 2010-02-25]. Faktorová analýza. Dostupné z WWW:
<http://www.spss.cz/cv/cv_fa.htm>.
- [14] SVOBODOVÁ, L. *Kvalita života* [online]. [s.l.], 2007. 18 s. Oborová práce. Výzkumný ústav bezpečnosti práce. Dostupné z WWW:
<http://kvalitazivota.vubp.cz/prispevky/soudoba_sociologie_II_kvalita_zivota-svobodova.doc>.
- [15] ÚBERLA, K. *Faktorová analýza*. 2. vyd. Bratislava : ALFA, 1976. 334 s.

Seznam příloh

Příloha 1 – Faktory, které vytvářejí kvalitu života	52
Příloha 2 – Model kvality života	52
Příloha 3 – Mapa světa s rozdělením HDI v členských státech OSN v roce 2007.....	53
Příloha 4 – Ukázka prvních patnácti záznamů datové matice	54
Příloha 5 – Stream: Metoda hlavních os, matice A	55
Příloha 6 – Stream: Metoda hlavních komponent, matice A.....	55
Příloha 7 – Stream: Metoda hlavních os, matice B	56
Příloha 8 – Stream: Metoda hlavních komponent, matice B.....	56

Příloha 1 – Faktory, které vytvářejí kvalitu života [zdroj: 1]

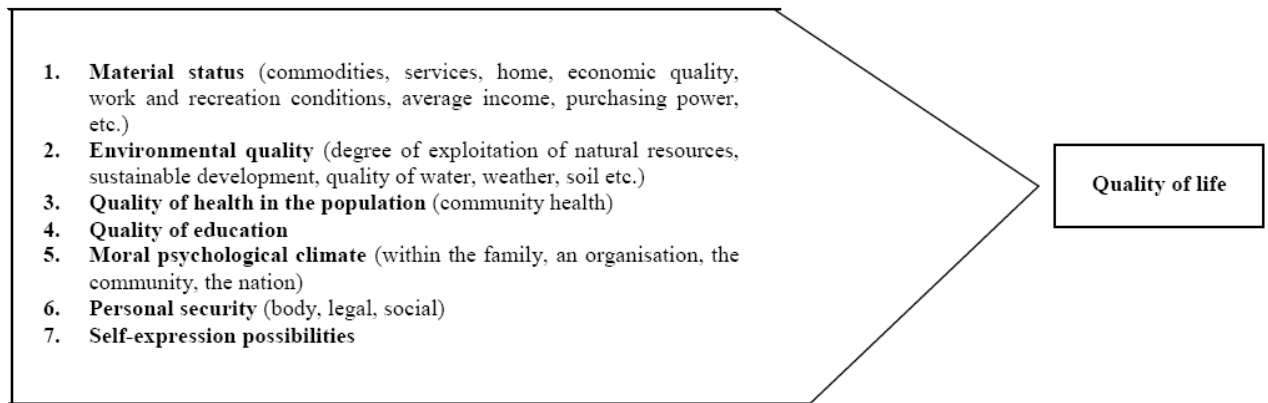


Figure 1. Factors which make up quality of life

Příloha 2 – Model kvality života [zdroj: 1]

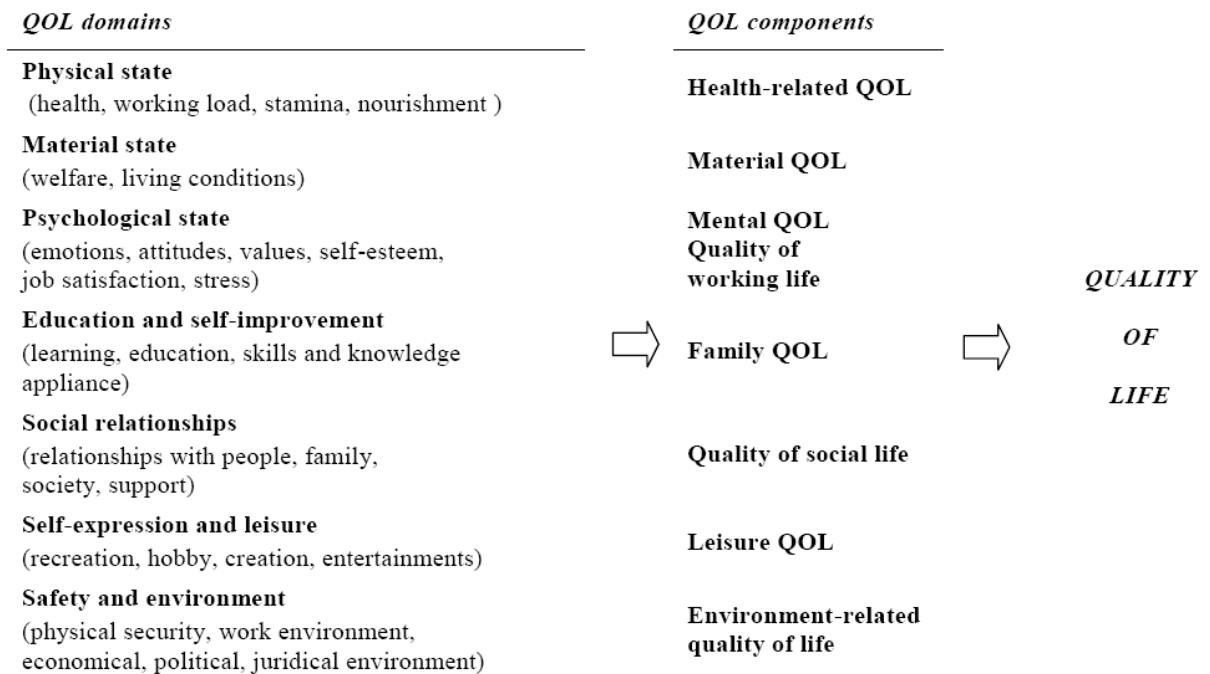
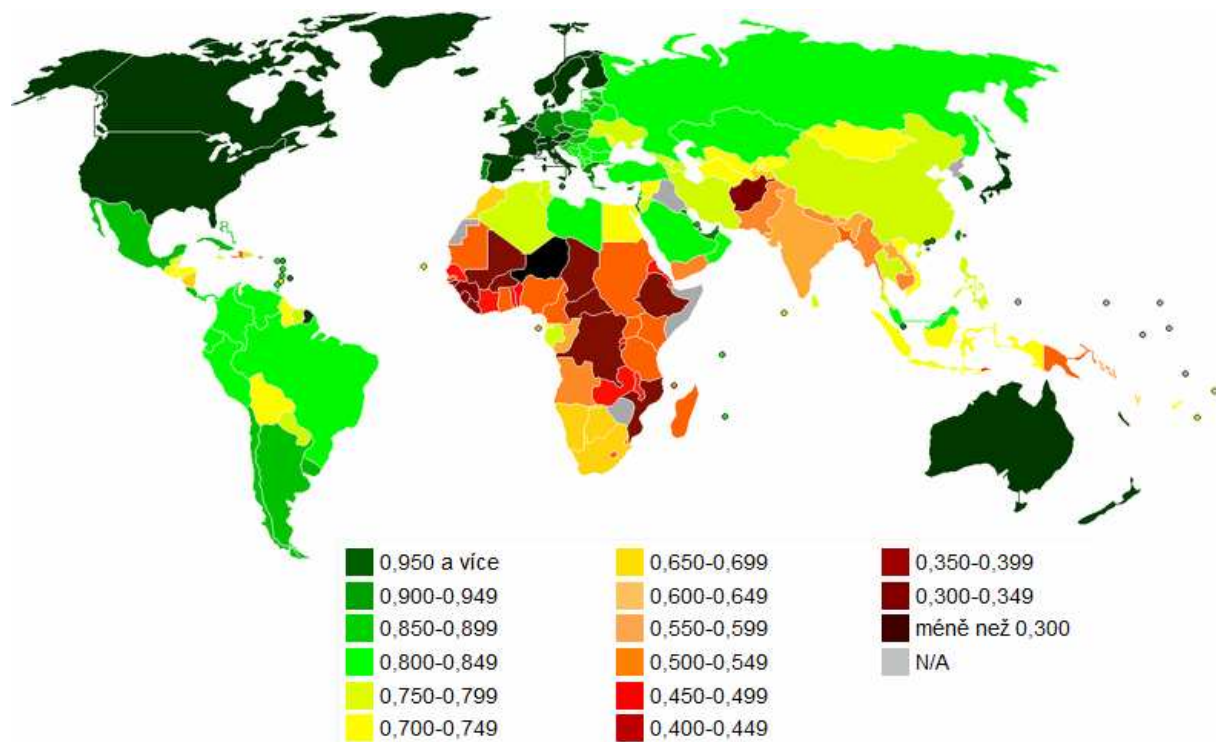


Figure 2. The quality of life model

Příloha 3 – Mapa světa s rozdělením HDI v členských státech OSN v roce 2007 [zdroj: 9]

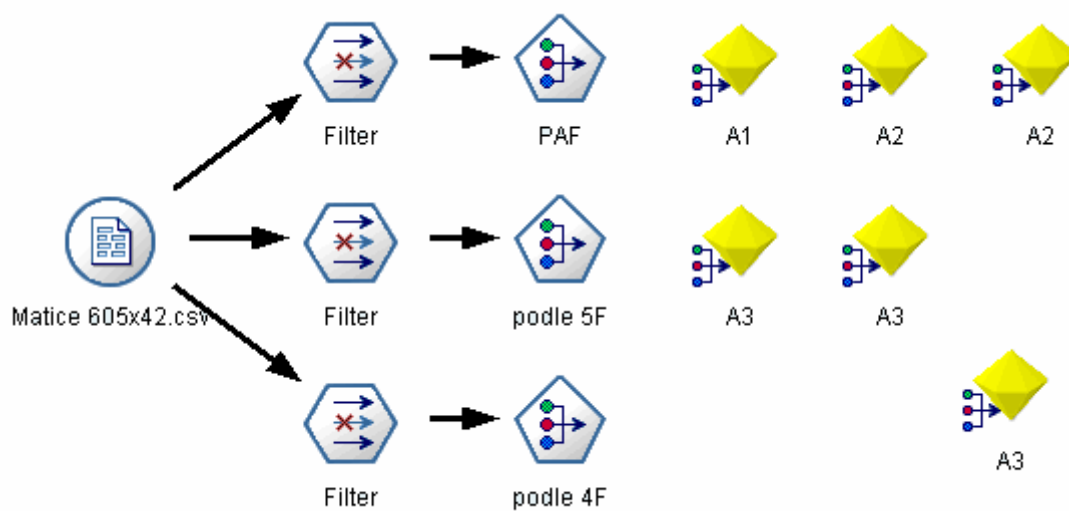


Příloha 4 – Ukázka prvních patnácti záznamů datové matice [zdroj: vlastní]

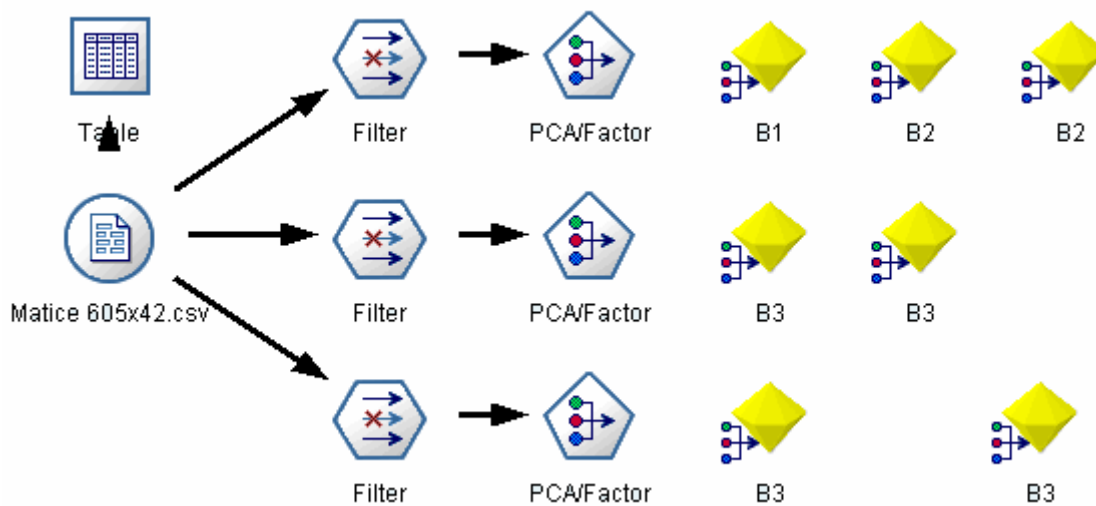
1			P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24
2	2001	Benešov	1524,0	61,0	39,0	62,3	3,0	88	6068	6149	326	164855	8,8	14,6	3,9	36,2	5,0	2,2	52,3	0,6	3,48	715	1243	3,2	324	1
3	2001	Beroun	662,0	114,0	39,4	53,0	2,1	103	6179	6215	2618	312749	8,3	14,7	8,2	48,7	5,3	2,7	55,2	1,5	4,96	483	697	3,1	322	2
4	2001	Kladno	692,0	218,0	39,1	70,1	1,5	55	6204	6397	260	246178	8,6	25,1	11,5	19,7	5,5	3,3	63,3	2,2	9,80	771	776	3,1	318	3
5	2001	Kolín	846,0	113,0	39,7	71,3	2,0	42	6137	6203	966	125890	5,0	16,3	7,5	23,4	5,7	2,5	51,6	-0,5	9,79	556	838	3,1	318	2
6	2001	Kutná Hora	917,0	80,0	39,3	66,1	2,3	91	6047	6157	266	97524	6,8	13,5	5,7	29,3	4,5	2,6	54,3	-3,9	11,31	189	894	2,6	355	2
7	2001	Mělník	712,0	134,0	38,6	66,1	1,6	51	6119	6216	2352	657264	9,8	97,4	81,6	31,5	5,4	3,7	61,6	0,6	7,39	369	612	2,9	349	2
8	2001	Mladá Boleslav	1058,0	106,0	38,5	62,9	2,3	102	6092	6228	977	4280993	5,4	16,2	8,3	26,9	6,2	3,5	42,3	1,6	3,17	1009	932	3,3	299	1
9	2001	Nymburk	876,0	94,0	39,5	69,9	1,7	60	6085	6199	2162	139992	6,0	13,3	4,4	30,1	5,9	3,1	42,5	4,0	8,86	765	749	2,5	377	2
10	2001	Praha-východ	584,0	162,0	39,5	70,3	4,9	63	6175	6297	243	256495	4,6	6,8	2,6	18,8	5,6	2,9	50,4	12,1	2,96	1682	674	2,4	406	3
11	2001	Praha-západ	586,0	138,0	39,3	59,4	8,0	107	6247	6386	479	250153	6,1	11,9	4,5	27,4	5,0	2,7	44,7	25,0	2,82	753	582	1,6	629	2
12	2001	Příbram	1628,0	66,0	38,6	43,6	2,7	162	6263	6425	335	141826	9,7	46,5	15,7	43,6	5,2	2,9	45,6	-1,0	8,33	777	905	3,5	274	5
13	2001	Rakovník	930,0	58,0	39,1	52,5	1,3	62	6092	6228	959	135622	8,4	16,0	5,0	37,8	5,3	3,0	47,5	-1,2	8,62	172	663	2,7	376	1
14	2001	České Budějovice	1625,0	110,0	38,5	53,2	1,9	335	6132	6278	1134	382197	4,3	18,5	7,0	16,5	5,9	2,9	56,5	1,4	4,43	500	1083	4,4	227	1
15	2001	Český Krumlov	1615,0	37,0	36,4	35,8	2,2	734	6059	6138	1680	37905	11,3	22,8	8,3	42,2	6,4	3,3	64,6	0,4	8,73	265	693	2,7	376	1

1			P25	P26	P27	P28	P29	P30	P31	P32	P33	P34	P35	P36	P37	P38	P39	P40	P41	P42
2	2001	Benešov	16	34079	25,81	6,166	28,2	25,4	166	10	117	8	4	0	4	10	117	32	14	2
3	2001	Beroun	17	29797	27,39	6,444	49,9	22,8	208	14	90	8	0	0	1	15	170	43	5	2
4	2001	Kladno	29	47085	27,72	6,318	36,8	18,5	383	4	105	8	1	7	4	14	160	81	8	3
5	2001	Kolín	16	31710	23,75	4,555	33,6	17,4	225	2	91	6	3	2	0	3	98	52	7	1
6	2001	Kutná Hora	13	27706	24,60	5,641	34,6	13,4	117	8	77	9	15	3	5	5	83	39	7	3
7	2001	Mělník	16	33026	23,37	4,794	44,5	18,5	311	13	103	11	4	0	4	10	119	51	11	3
8	2001	Mladá Boleslav	17	61552	24,34	5,467	51,6	28,3	234	9	135	6	2	2	3	18	109	43	5	8
9	2001	Nymburk	14	28167	25,64	5,013	33,6	21,0	149	5	90	9	4	2	3	6	114	45	12	2
10	2001	Praha-východ	16	38268	28,23	5,968	55,2	30,0	284	7	93	6	5	0	3	16	134	55	11	2
11	2001	Praha-západ	14	30855	28,82	5,699	47,6	36,7	258	10	70	5	4	0	4	4	101	45	9	3
12	2001	Příbram	19	38700	28,03	6,381	38,6	22,2	303	4	105	8	3	3	5	8	96	55	8	4
13	2001	Rakovník	9	19331	29,06	6,619	33,9	20,2	108	7	89	4	4	3	5	7	105	43	5	2
14	2001	České Budějovice	29	105014	26,42	6,242	37,4	27,2	353	12	167	12	30	6	4	28	314	81	12	6
15	2001	Český Krumlov	12	21992	23,46	5,998	25,9	24,5	176	3	52	6	16	1	6	28	74	48	10	1

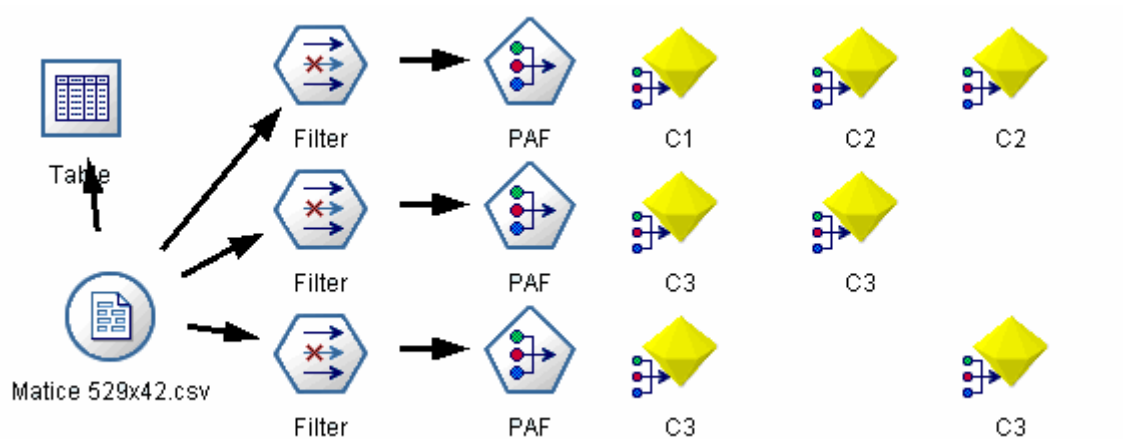
Příloha 5 – Stream: Metoda hlavních os, matice A [zdroj: vlastní]



Příloha 6 – Stream: Metoda hlavních komponent, matice A [zdroj: vlastní]



Příloha 7 – Stream: Metoda hlavních os, matice B [zdroj: vlastní]



Příloha 8 – Stream: Metoda hlavních komponent, matice B [zdroj: vlastní]

