

ZX510 Pokročilé statistické metody geografického výzkumu

Téma: Explorační faktorová analýza
(analýza hlavních komponent)



evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

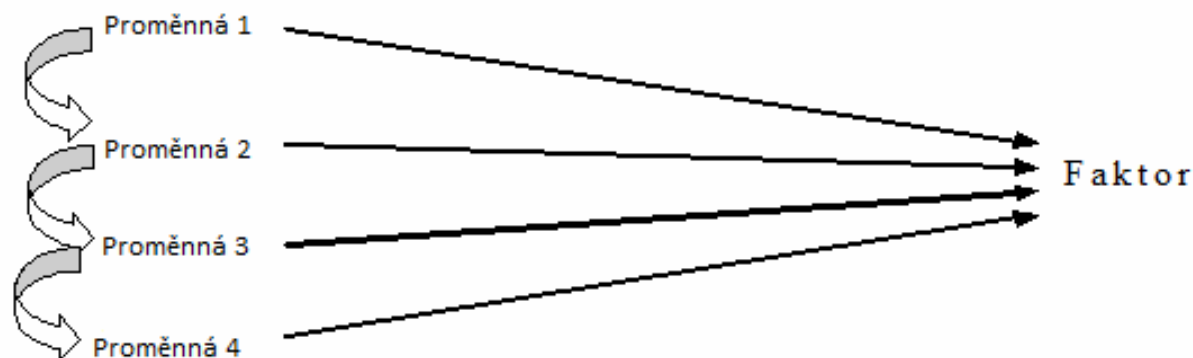


OP Vzdělávání
pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Smysl a princip faktorové analýzy



- v explorační verzi není faktorová analýza určena k meritorní analýze, tzn. neslouží k testování hypotéz ani k měření vztahů mezi nezávisle a závisle proměnnou
- jejím cílem je především redukce dat a odhalení (explorace) skrytých souvislostí
- z mnoha položek (proměnných) vybere ty, které statisticky „patří k sobě“ (vykazují podobnou variabilitu) a z nichž je možné vytvořit novou proměnnou (faktor, komponentu)



Příklad: analýza faktorů rozvojového potenciálu obcí Jihomoravského kraje

- Datový soubor: [Faktorovka_priklad.sav](#)
- Obsahuje data za obce JMK – u kterých předpokládáme, že představují dílčí indikátory rozvojového potenciálu obcí (např. vzdálenost od regionální centra, počet obyvatel, míra nezaměstnanosti, bytová výstavba, index stárí, míra vzdělanosti, ekonomická aktivita, infrastruktura, daňový výtěžnost, atd.)
- Cílem faktorové analýzy je zjistit, zda proměnné společně tvoří určité komponenty, představující faktory rozvojového potenciálu obcí

Faktorová analýza v SPSS

Faktorovka_priklad.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform **Analyze** Direct Marketing Graphs Utilities Add-ons Window Help

Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
Multiple Imputation
Complex Samples
Quality Control
ROC Curve...

byty obyvatele hu

Factor...
Correspondence Analysis...
Optimal Scaling...

	code	vzdal_O	byty	obyvatele	hu
1	582786,00		15,00	378965,00	
2	584801,00		29,25	4640,00	
3	581283,00		23,19	20769,00	
4	584649,00		40,12	7374,00	
5	586021,00		2,84	25259,00	
6	593711,00		17,87	34073,00	
7	595098,00	2	8,17	860,00	
8	594482,00		4,67	5903,00	
9	581879,00		10,16	2720,00	
10	582239,00				
11	584291,00				
12	586811,00				
13	582158,00	2	6,29	1722,00	
14	583898,00		71,84	2220,00	
15	584002,00		38,64	8714,00	
16	586307,00		10,02	11548,00	
17	586803,00		11,33	2603,00	
18	593117,00		19,56	2905,00	
19	593770,00	2	7,31	1218,00	
20	595110,00	4	31,41	294,00	



Postup faktorové analýzy

- (1) Typ dat - musíme rozhodnout, zda jsou data pro faktorovou analýzu vhodná;
- (2) Počet faktorů - musíme rozhodnout, s jakým počtem faktorů (komponent) budeme pracovat;
- (3) Výpočet sycení faktorů a jejich pojmenování - musíme vypočítat faktorové zátěže a jednotlivé faktory pojmenovat



Ad 1. Vhodnost dat pro faktorovou analýzu

- Položky vstupující do faktorové analýzy, musí mít podobu ordinálních nebo intervalových proměnných, nevhodné jsou dichotomické proměnné (výjimečně je možné použít tzv. *dummy variables*)
- Proměnné by mezi sebou měly korelovat (viz. výstup v podobě tzv. matice interkorelací), ale neměl by mezi nimi být kauzální vztah
- Míru vhodnosti dat zjistíme také s tzv. anti-image matice s hodnotami *Kaiser-Meyer-Olkinovy míry* - KMO míra vhodnosti položek pro faktorovou analýzu. Hodnota KMO by měla být vyšší než 0,6.
- Signifikance tzv. *Bartlettova testu sféricity* by měla být významná minimálně na úrovni 0,05.

Vhodnost dat pro faktorovou analýzu

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is highlighted with a red circle. A red arrow points from the 'Analyze' menu to the 'Factor Analysis' button in the 'Factor Analysis' dialog box. The 'Factor Analysis' dialog box is open, showing a list of variables on the left and a list of variables in the 'Variables:' box on the right. The 'Descriptives...' button is highlighted with a red circle. A red arrow points from the 'Descriptives...' button to the 'Factor Analysis: Descriptives' dialog box. The 'Factor Analysis: Descriptives' dialog box is open, showing the 'Statistics' section with 'Univariate descriptives' unchecked and 'Initial solution' checked. The 'Correlation Matrix' section has 'Coefficients' unchecked, 'Inverse' unchecked, 'Significance levels' checked, 'Reproduced' unchecked, 'Determinant' checked, 'Anti-image' checked, and 'KMO and Bartlett's test of sphericity' checked. The 'Continue' button is highlighted with a red circle.

	code	vzdal_ORP	vzdal_Brno	zeleznice	byty	obyvatele	hustota	přirustek
1						78965,00	1646,24	2,03
2						4640,00	107,78	9,83
3						20769,00	461,84	2,31
4						7874,00	162,71	,22
5						25259,00	398,97	-8,43
6						44073,00	517,28	-2,50
7						860,00	64,08	-7,15
8						5903,00	119,11	-1,99
9						2739,00	113,18	12,66
10						3545,00	232,76	4,64
11						27015,00	324,41	-3,03
12								
13								
14								
15	584002,00	,00	25,00	1,00	38,1			
16	586307,00	,00	54,00	1,00	10,1			
17	586803,00	14,00	44,00	,00	11,1			
18	593117,00	8,30	44,00	1,00	19,1			
19	593770,00	25,00	79,00	1,00	7,1			
20	595110,00	45,00	105,00	,00	31,1			
21	581372,00	,00	39,00	1,00	26,1			
22	546941,00	3,40	63,00	,00	13,1			
23	583588,00	4,60	29,00	1,00	14,1			
24	584550,00	18,00	34,00	,00	25,1			



Ad 2. Extrakce faktorů

Musíme rozhodnout, s kolika faktory máme pracovat

- výpočet tzv. eigen hodnoty (*eigenvalue*) volíme takový počet faktorů (komponent), které mají hodnotu eigenvalue vyšší než 1 (tzv. *Kaiserovo pravidlo*)
- zvolíme takový počet faktorů, který nám bude vysvětlovat dostatečné % variability v souboru
- zvolíme takový počet faktorů, které v grafu (*Scree plot*) budou nad tzv. zlomem či prolomením křivky

Ad 2. Extrakce faktorů

Faktorovka_priklad.sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help

code vzdal_ORP vzdal_Brno zeleznice byty obyvatele hustota prirustek index_stari vzdal

1 378965,00 1646,24 2,03 136,50

2 4640,00 107,78 9,83 109,50

3 20769,00 461,84 2,31 128,40

4 7374,00 162,71 ,22 101,00

5 25259,00 398,97 -8,43 120,30

6 34073,00 517,28 -2,50 114,60

7 860,00 64,08 -7,15 126,70

8 5503,00 119,11 -1,99 119,60

9 2735,00 113,18 12,66 77,40

10 3645,00 232,76 4,64 111,00

11

12

13

14

15 584002,00 ,00 25,00 1,00

16 586307,00 ,00 54,00 1,00

17 586803,00 14,00 44,00 ,00

18 593117,00 8,30 44,00 1,00

19 593770,00 25,00 79,00 1,00

20 595110,00 45,00 105,00 ,00

21 581372,00 ,00 39,00 1,00

22 546941,00 3,40 63,00 ,00

Factor Analysis

Variables:

code

faktor_perifernost

faktor_rust

faktor_podnikatel

faktor_vybavenost

faktor_celkem

vzdal_ORP

vzdal_Brno

zeleznice

byty

obyvatele

hustota

prirustek

Selection Variable:

Value

OK Paste Reset Cancel Help

Descriptives...

Extraction...

Rotation...

Scores...

Options...

Factor Analysis: Extraction

Method: Principal components

Analyze

☒ Correlation matrix

☐ Covariance matrix

Display

☐ Unrotated factor solution

☒ Scree plot

Extract

☒ Based on Eigenvalue

Eigenvalues greater than: 1

☐ Fixed number of factors

Factors to extract:



KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.			,735
Bartlett's Test of Sphericity	Approx. Chi-Square		826,112
	df		91
	Sig.		,000

Výstupy z faktorové analýzy

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings ^a
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	3,858	27,554	27,554	3,858	27,554	27,554	2,426
2	2,559	18,282	45,836	2,559	18,282	45,836	2,199
3	1,685	12,034	57,870	1,685	12,034	57,870	2,826
4	1,232	8,797	66,667	1,232	8,797	66,667	2,985
5	,770	5,503	72,170				
6	,699	4,993	77,163				
7	,613	4,376	81,539				
8	,531	3,794	85,333				
9	,521	3,718	89,051				
10	,462	3,299	92,350				
11	,361	2,578	94,929				
12	,273	1,947	96,876				
13	,249	1,781	98,657				
14	,188	1,343	100,000				

Výstupy z faktorové analýzy





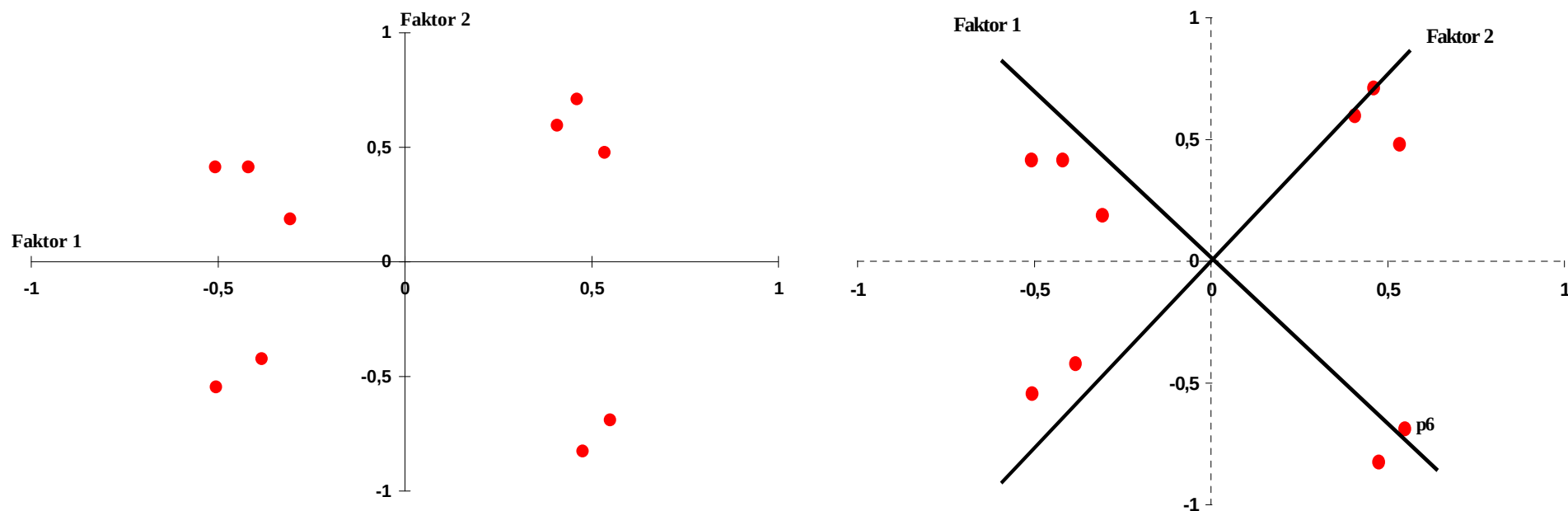
Ad 3. Výpočet faktorových zátěží a pojmenování faktorů

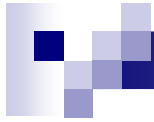
- výpočet faktorových zátěží (*factor loadings*)
 - tyto zátěže jsou korelace mezi faktorem a příslušnou položkou
 - čím vyšší je korelace, tím více je faktor touto položkou sycen (korelace by měla být vyšší než 0,30).

Rotace faktorů

- smyslem rotace faktorů je, aby se původně rozptýlené body co nejvíce přimkly k jednomu z extrahovaných faktorů

Obr. rozptyl hodnot kolem faktorů při nerotovaném a rotovaném řešení





Základní typy rotace

- Rotace **Varimax** – ortogonální (pravoúhlá, kolmá)
 - extrahované faktory nemusí korelovat
 - výstupem je tabulka tzv. ***pattern matrix*** ve které hodnoty představují korelace a zároveň sycení (*loadings*) mezi faktory a proměnnými



Základní typy rotace

- Rotace **Oblimin** – šikmá
 - předpokládáme, že extrahované faktory navzájem korelují !
 - Výstupem jsou dvě tabulky: tzv. ***pattern matrix***
 - ve které hodnoty sycení (*loadings*) představují regresní koeficient, který určuje hodnotu proměnné
 - a tzv. ***structure matrix*** – představuje hodnoty korelace mezi faktorem a proměnnou

Výstupy – nerotované a nestrukturované řešení

Component Matrix^a

	Component			
	1	2	3	4
vzdal_ORP	-,651	,113	,025	,291
vzdal_Brno	-,548	,505	-,255	,398
zeleznice	,573	,231	-,357	,185
byty	,240	-,636	,410	,440
hustota	,707	,227	,187	-,070
přirustek	,207	-,777	,292	,306
index_stari	-,257	,489	,309	-,443
vzdelanost	,795	,225	,280	-,039
neuamestannost	-,604	,337	-,024	,466
EA_tercier	,599	,427	,133	,172
danova_vyteznost	-,020	,573	,511	,278
podnik_aktivita	,277	,473	,500	,130
vybavenost	,561	,195	-,590	,055
infrastruktura	,625	,059	-,390	,346

Extraction Method: Principal Component Analysis.

Výstupy – nerotované strukturované řešení

Component Matrix^a

	Component			
	1	2	3	4
vzdelanost	,795			
hustota	,707			
vzdal_ORP	-,651			
infrastruktura	,625		-,390	,346
neuamestannost	-,604	,337		,466
EA_tercier	,599	,427		
zeleznice	,573		-,357	
vzdal_Brno	-,548	,505		,398
přirustek		-,777		,306
byty		-,636	,410	,440
danova_vyteznost		,573	,511	
index_stari		,489	,309	-,443
vybavenost	,561		-,590	
podnik_aktivita		,473	,500	

Extraction Method: Principal Component Analysis.

Výstupy – rotované řešení - Varimax

Rotated Component Matrix^a

	Component			
	1	2	3	4
vybavenost	,798			
infrastruktura	,789			
zeleznice	,708			
vzdal_Brno		,824		-,307
neuamestannost		,817		
vzdal_ORP		,652		
podnik_aktivita			,751	
danova_vyteznost			,743	
vzdelanost	,348	-,519	,607	
EA_tercier	,438		,607	
hustota	,343	-,468	,505	
byty				,891
přirustek				,868
index_stari	-,432			-,567

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

Výstupy – rotované řešení - Oblimin

Pattern Matrix^a

	Component			
	1	2	3	4
danova_vyteznost	,774			,327
podnik_aktivita	,759			
EA_tercier	,584		-,387	
vzdelanost	,574			-,466
hustota	,468			-,424
byty		-,906		
přirustek		-,858		
index_stari		,550	,490	
infrastruktura			-,800	
vybavenost			-,792	
zeleznice			-,700	
vzdal_Brno				,850
neuamestannost				,841
vzdal_ORP				,642

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

Pojmenování faktorů u vybraného řešení (Oblimin)

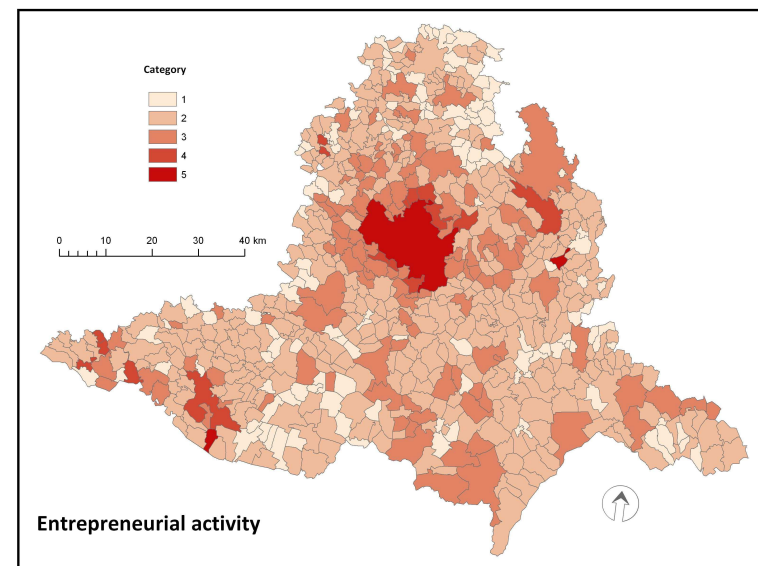
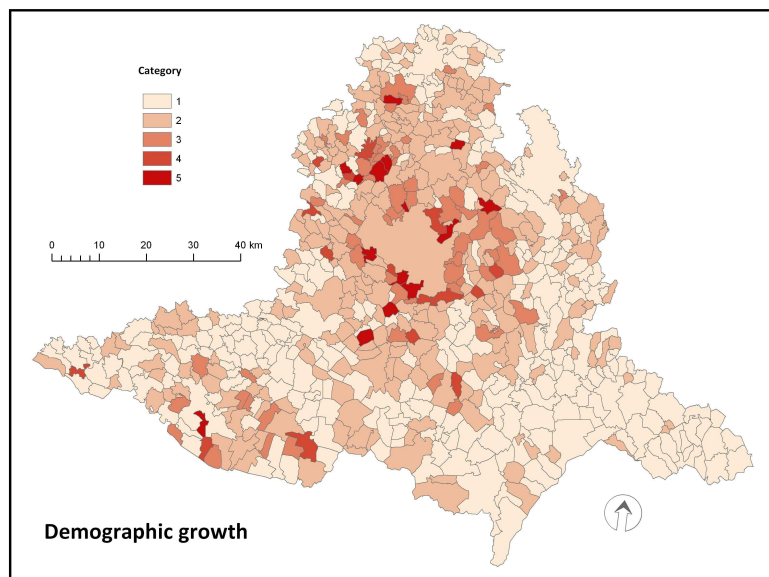
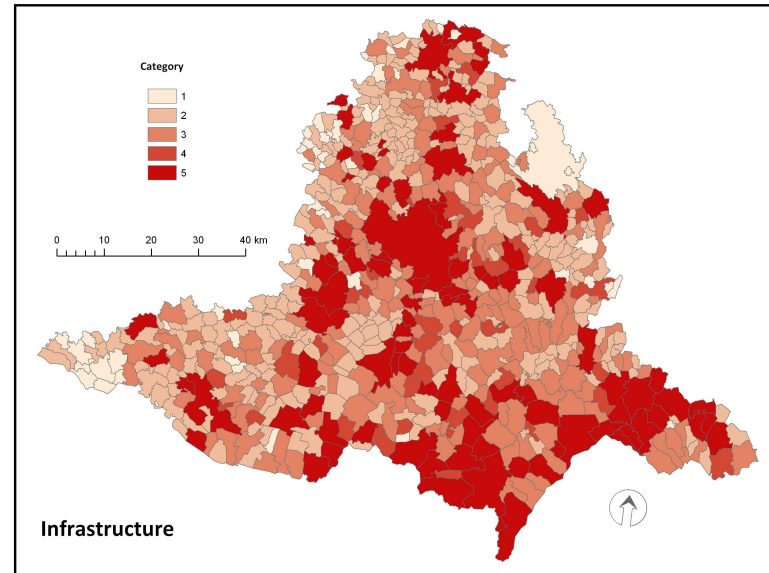
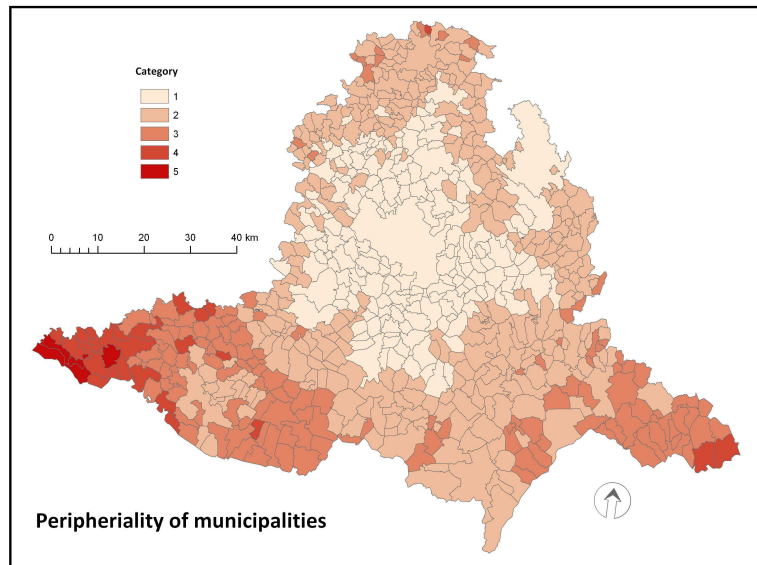
Pattern Matrix^a

	Component			
	1 „ekonomický potenciál“	2 „demografický růst“	3 „vybavenost“	4 „perifernost“
daňová výtěžnost	,774			,327
podnikatelská_aktivita	,759			
EA_terciér	,584		-,387	
vzdělanost	,574			-,466
hustota	,468			-,424
dokončené byty		,906		
přírůstek obyvatel		,858		
index_stari		-,550	,490	
technická infrastruktura			-,800	
občanská vybavenost			-,792	
zeleznice			-,700	
vzdálenost od Brno				,850
nezaměstnanost				,841
vzdálenost od ORP				,642

Extraction Method: Principal Component Analysis.

Rotation Method: Oblimin with Kaiser Normalization.

Ukázka mapového zpracování výsledků analýzy – kategorizace obcí dle hodnot dílčích faktorů



Ukázka mapového zpracování výsledků analýzy
– kategorizace obcí dle celkové hodnoty rozvojového potenciálu

