FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

UNIVERZITY KOMENSKÉHO

v Bratislave

DIPLOMOVÁ PRÁCA

2001 **Juraj Holec**

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY UNIVERZITY KOMENSKÉHO v Bratislave

Priestorové a štrukturálne analýzy nezamestnanosti v SR

Diplomová práca

Diplomant: Juraj Holec

Vedúci diplomovej práce: Mgr. Gabriela Moravčíková Bratislava 2001

Čestné vyhlásenie

Prehlasujem , že na diplomovej práci som pracoval samostatne na základe vlastných teoretických a praktických poznatkov, konzultácií a štúdia odbornej literatúry, ktorej úplný prehľad som uviedol v zozname.

V Bratislave, dňa 30.3.2001

Poďakovanie

Týmto by som chcel poďakovať svojej vedúcej diplomovej práce, Mgr. Gabriele Moravčíkovej, za odborné rady a pomoc pri tvorbe diplomovej práce

Obsah

Úvod

Analýza údajov o trvaní nezamestnanosti poskytuje zásadné podklady pre politiku trhu práce, pre rozhodovanie o otázkach výšky príspevkov v nezamestnanosti, dobe poberania príspevkov v nezamestnanosti, o rekvalifikačných programoch a podobne.

Samotná existencia nezamestnanosti je prirodzeným fenoménom a atribútom slobodnej spoločnosti založenej na trhovom hospodárstve a demokracii. Jej nekontrolovateľný vývoj spôsobujúci masový charakter však vyvoláva nielen vážne ekonomické, ale aj sociálne problémy (rozpad rodiny, narušené mentálne i fyzické zdravie, trestná činnosť a ďalšie sociálno-patologické javy). Skutočnosť, že súčasná situácia na trhu práce je poznačená nerovnováhou, ktorá má frikčný charakter len v minimálnom rozsahu a veľká časť nezamestnanosti má charakter nepravej cyklickej, čiastočne aj štrukturálnej a sezónnej nezamestnanosti, signalizuje, že riešenie problému nezamestnanosti presahuje rámec opatrení, nástrojov a programov politiky trhu práce. Hlavnou príčinou nezamestnanosti v SR je pokles dynamiky rastu a následný úpadok slovenského hospodárstva.

Prvý krát bol prudký nárast nezamestnanosti zaznamenaný v SR v roku 1991. Jeho príčinou bola najmä konverzia zbrojnej výroby, zdraženie vstupov z dovozu v surovinovo a materiálovo náročných výrobách a rozpad východných trhov. Prudký nárast nezamestnanosti zaznamenaný ku koncu roka 1998 bol ovplyvnený predovšetkým rastúcimi zdrojmi pracovných síl v dôsledku demografického vývoja a poklesom zamestnanosti spôsobeným znížením dynamiky ekonomického rastu z makroekonomického hľadiska, nedostatkom finančných zdrojov v podnikoch a s tým súvisiacim úpadkom a bankrotmi podnikov, hromadným prepúšťaním zamestnancov. Významnú úlohu zohrali najmä systémové dôsledky odsunutia bankrotov podnikov v predchádzajúcom období, čo sa však prejavilo v raste nezamestnanosti na prelome rokov 1998/1999. Hromadným prepúšťaním k 31. 12. 1998 v zmysle § 112 zákona NR SR č. 387/1996 Z. z. o zamestnanosti bolo napríklad ohrozených 24 889 až 25 109 pracovných miest (k 30. 6. 1999 to bolo 11 277 ohrozených pracovných miest).

Vývoj nezamestnanosti v SR je zároveň ovplyvnený nedostatočnou väzbou medzi školským systémom a trhom práce, nízkou mobilitou pracovnej sily, nevhodným pomerom medzi sociálnymi a pracovnými príjmami, nízkou tvorbou pracovných miest, obmedzenými možnosťami použitia nástrojov aktívnej politiky trhu práce a nejasnými vlastníckymi vzťahmi najmä v bankrotujúcich podnikoch. Nezdravé tendencie v spoločnosti, prameniace najmä zo zmeny vlastníckych vzťahov a ktoré významne narušili princíp sociálnej spravodlivosti, znižujú hodnotu práce nielen ako zdroja obživy, ale aj ako prostriedku sebaúcty a spoločenského uplatnenia občanov.

Kategorizácia nezamestnanosti na frikčnú, štrukturálnu a cyklickú vznikla v podmienkach dlhoročne sa rozvíjajúcich trhových ekonomík. Nezamestnanosť v SR, ako v krajine, ktorej ekonomika prechádza transformačným procesom, má svoje špecifiká. Samotný pojem cyklická alebo konjunkturálna nezamestnanosť nevystihuje celkom situáciu v ekonomike SR, kde neboli zatiaľ identifikované typické hospodárske cykly. Nezamestnanosť v SR sa však vyznačuje nepochybne aj znakmi, ktoré sú obsiahnuté v definícii cyklickej (konjunkturálnej) nezamestnanosti (celkový vysoký počet evidovaných nezamestnaných pripadajúcich na jedno voľné pracovné miesto, celkový nízky dopyt po pracovných silách spôsobený najmä značným poklesom zamestnanosti v niektorých odvetviach priemyslu, v pôdohospodárstve, stavebníctve, bez toho, že by sa adekvátne vytvárali nové pracovné miesta), ale aj znakmi štrukturálnej nezamestnanosti. Štrukturálna nezamestnanosť znamená nesúlad medzi profesijnou a/alebo územnou štruktúrou voľných pracovných miest a profesijnou a územnou štruktúrou nedobrovoľne nezamestnaných. Štrukturálna nezamestnanosť býva často dôsledkom dynamickej reštrukturalizácie ekonomiky, keď so zánikom neefektívnych pracovných miest a znižovaním prezamestnanosti v niektorých odvetviach vznikajú nové pracovné miesta a ich obsadeniu nezamestnanými bráni len nesúlad medzi kvalifikačnou a územnou štruktúrou nezamestnaných a štruktúrou voľných pracovných miest. Na odstránenie štrukturálnej nezamestnanosti sa používajú programy politiky trhu práce podporujúce profesijnú a územnú mobilitu pracovných síl (rekvalifikačné programy, zlepšenie informovanosti o voľných pracovných miestach, príspevky na prepravu do práce a späť a pod.). Cyklická alebo konjunkturálna nezamestnanosť sa vyskytuje vtedy, keď celkový dopyt po pracovníkoch je nízky (teda nielen dopyt po pracovníkoch v určitých oblastiach alebo v určitých profesiách) a keď zamestnanosť klesá v dôsledku nedostatočného agregálneho dopytu, t.j. celkového súhrnu všetkých plánovaných alebo želaných výdavkov v ekonomike v danom období. Tento typ nezamestnanosti sa vyznačuje nedostatkom voľných pracovných miest. Pri riešení nezamestnanosti cyklického typu zohrávajú nástroje politiky trhu práce skôr doplnkovú úlohu a ťažisko politiky zamestnanosti sa presúva do oblasti hospodárskej politiky, ktorá by mala vytvoriť podmienky pre dynamický ekonomický rast vytvárajúci pracovné príležitosti.

Cieľom tejto práce je naštudovať a aplikovať vybrané metódy vhodné na porovnávanie územných jednotiek Slovenska z pohľadu nezamestnanosti. Existuje niekoľko ekonomických ukazovateľov, ktoré spoločne vytvárajú obraz o situácii na trhu práce. Preto porovnávanie založené len na jednom ukazovateli by nemuselo byť úplné. Cieľom tejto diplomovej práce je naštudovať vybrané štatistické metódy, ktoré uľahčujú proces porovnávania okresov a pritom berú do úvahy viacero štatistických znakov súčasne. Metódy ktorým sa budem venovať sú metódy viackriteriálneho porovnávania, zhluková a faktorová metóda. Najprv ich popíšeme a zosumarizujeme ich možnosti aplikácii. Následne ich budeme aplikovať. Pri aplikácii nás bude zaujímať aj stabilnosť riešenia v čase (porovnanie okresov robíme dva roky po sebe) a porovnateľnosť výsledkov medzi metódami.

V prvej kapitole spomenieme situáciu na trhu práce, popíšeme jednotlivé ukazovatele (štatistické, ukazovatele nezamestnanosti), a spomenieme metódy viackriteriálneho porovnávania.

V druhej kapitole sa budeme zaoberať popísaním metód zhlukovej a faktorovej analýzy.

Kapitola číslo tri sa bude venovať jednotlivým výsledkom a interpretáciám výsledkov analýz za sledované mesiace. Na analyzovanie pomocou metód viackriteriálneho porovnávania som používal program Microsoft Excel a na metódy faktorovej a zhlukovej analýzy som použil program S+.

Kapitola 1

Situácia na trhu[6]

Na porovnávanie som si vybral mesiac december v rokoch 1999 a 2000.

Neustály pokles celkovej nezamestnanosti v hospodárstve SR, ako aj definitívne ukončenie sezónnych prác, výrazne ovplyvňovali vývoj situácie v nezamestnanosti v decembri roku 1999. Prítok do evidencie úradov práce za posledné dva mesiace je historicky najvyšší v existencii samostatnej SR. V decembri roku 1999 dominuje takmer 10 tis. novoevidovaných nezamestnaných osôb zo stavebníctva. Priemysel si „udržiava“ stabilný prítok v rozmedzí 5 – 7 tis. osôb.

Napriek týmto nepriaznivým trendom sme svedkom i jedného priaznivého trendu. Už štvrtý mesiac po sebe je zaznamenaný rast celkového ročného počtu vyradených nezamestnaných z evidencie úradu práce, keď tento v decembri prekročil hranicu 355 tis. osôb. Tento výsledok je prejavom neustále sa zintenzívňujúcej činnosti národného úradu práce (NÚP) v oblasti sprostredkovateľských a poradenských služieb. Nestačí však eliminovať neustály rast prílevu nezamestnaných do evidencie úradov práce (ÚP). Aj keď výsledkom toho je zníženie medziročného rozdielu v stave nezamestnaných na (v porovnaní s novembrom) na hodnotu 107 tis. osôb, počet nezamestnaných ku koncu roku 1999 oproti novembru 1999 vzrástol o viac ako 21 tis. osôb, čím prekročil hranicu 535 tis. osôb.

Vzhľadom k tomu, že zamestnanosť v hospodárstve SR nevykazuje žiadne známky oživenia (a ani konjunkturálny prieskum ŠÚ SR nenaznačuje zmenu v najbližších mesiacoch), NÚP očakáva koncom januára počet nezamestnaných na úrovni 560 tis. osôb.

Stav evidovaných nezamestnaných k 31.12.2000 dosiahol 506497 osôb. V porovnaní s decembrom 1999 stav evidovaných nezamestnaných ku koncu decembra 2000 poklesol o 28714 osôb (o –5,36%). Miera nezamestnanosti disponibilných evidovaných nezamestnaných medziročne poklesla o 1,3 percentuálneho bodu na hodnotu 17,88%.

Tento pozitívny výsledok, t.j. zníženie stavu evidovaných nezamestnaných ku koncu roka 2000 je dôsledkom viacerých faktorov, najmä však verejnoprospešných prác, legislatívnych zmien zákona o zamestnanosti v priebehu roka 2000 a nesporne tiež zvýšeným úsilím všetkých zložiek NÚP, ktoré sa prejavili v sprostredkovateľskej činnosti.

Hoci vplyv verejnoprospešných prác na zníženie nezamestnanosti je vysoký, jeho trvácnosť možno charakterizovať ako najkratšiu. Ako ukazuje decembrový prítok osôb do evidencie nezamestnaných, v mesiaci december 2000 bolo zaevidovaných o takmer 10 tisíc osôb viacej ako v decembri 1999. Tento medziročný nárast možno „pripísať najmä na konto" osobám „vracajúcim" sa z verejnoprospešných prác, ktorým skončila pracovná zmluva na konci roka 2000.

Verejnoprospešnými prácami sa pozastavil tiež ďalší prudký nárast počtu dlhodoboevidovaných nezamestnaných, pričom priemerný mesačný počet dlhodobo evidovaných nezamestnaných sa stabilizoval na hladine 240 tisíc a priemerný podiel na celkovom počte evidovaných nezamestnaných sa ustálil na úrovni 45%. Priemerná dĺžka evidencie evidovaných nezamestnaných tak ku koncu roka 2000 v porovnaní s rokom 1999 poklesla o 0,63 mesiaca na 13,73 mesiacov.

V priebehu mesiaca december 2000 bolo nahlásených celkovo iba 5616 voľných pracovných miest, pričom stav voľných pracovných miest ku koncu decembra 2000 dosiahol 6026 miest.

Z evidencie bolo vyradených 17518 osôb, z ktorých bolo 66% umiestnených do pracovného pomeru.

Kľúčové pojmy[6]

Evidovaný nezamestnaný	je občan hľadajúci zamestnanie, zaradený do evidencie nezamestnaných na OÚP po podaní písomnej žiadosti o sprostredkovanie zamestnania.
Disponibilný EN	je to EN, ktorý bezprostredne môže nastúpiť do pracovného pomeru.

Miera nezamestnanosti	podiel počtu EN, resp. disponibilného počtu EN k ekonomicky aktívnemu obyvateľstvu vyjadrený v percentách.
Prítok	súčet denných počtov osôb vstupujúcich do evidencií OÚP v priebehu sledovaného mesiaca. -súčet denných počtov osôb poberajúcich PvN v evidencii OÚP v priebehu sledovaného mesiaca. -súčet denných počtov PM, ktoré v priebehu sledovaného mesiaca nahlásia zamestnávatelia OÚP.
Odtok	súčet denných počtov osôb vyradených z evidencie OÚP v priebehu sledovaného mesiaca. -súčet denných počtov osôb vyradených z evidencie poberateľov PvN v priebehu sledovaného mesiaca. -súčet denných počtov PM, ktoré v priebehu sledovaného mesiaca OÚP prestal evidovať.
Priemerná dĺžka evidencie	vyjadruje čas (v mesiacoch), ako dlho je EN v priemere v evidencii OÚP.
Poberateľ (PvN)	EN, ktorý poberá PvN v súlade so zákonom o zamestnanosti.
Priemerná mesačná podpora	suma v Sk (bez odvodov do poistných fondov) vyplatená v priemere jednému EN v sledovanom mesiaci.
Voľné pracovné miesto	sa rozumie novovytvorené alebo existujúce neobsadené PM, na ktoré chce zamestnávateľ prijať zamestnancov.
EN mladistvý	je občan od nadobudnutia pracovnoprávnej spôsobilosti do dosiahnutia 18 rokov veku.
EN absolvent školy	je občan po dobu jedného roka od skončenia sústavnej prípravy na povolanie.

Ukazovatele nezamestnanosti[1]

(indexy)

UR	miera nezamestnanosti
RO	podiel vyradených uchádzačov z evidencie k priemernému počtu uchádzačov o zamestnanie
RI	podiel novozaregistrovaných uchádzačov k celkovému počtu zaregistrovaných uchádzačov
RUV	počet uchádzačov o zamestnanie pripadajúcich na jedno voľné pracovné miesto
EO	podiel vyradených uchádzačov umiestnených úradmi práce k celkovému počtu vyradených
LU13-48	podiel nezamestnaných viac ako rok a menej ako štyri roky k celkovému počtu dlhodobo nezamestnaných
LU48	podiel nezamestnaných viac ako štyri roky k celkovému počtu dlhodobo nezamestnaných
D	priemerná doba evidencie uchádzača o zamestnanie (podiel celkového počtu zaregistrovaných k počtu novozaregistrovaných uchádzačov)

Viacrozmerné priestorové porovnávanie stavu nezamestnanosti

Pre priestorové a štrukturálne porovnávanie rozsahu nezamestnanosti sa v praxi najčastejšie využíva ukazovateľ *miera nezamestnanosti.* Ak v určitom období chceme porovnať stav nezamestnanosti v okresoch SR, obyčajne ich zoradíme podľa hodnôt práve tohoto ukazovateľa a za najproblematickejší považujeme ten okres, kde je miera nezamestnanosti najvyššia. Rovnako zvýšenie hodnoty miery nezamestnanosti v porovnaní s predchádzajúcimi obdobiami považujeme jednoznačne za zhoršenie stavu nezamestnanosti.

Objektívnejšie je priestorové a štrukturálne porovnávanie stavu nezamestnanosti podľa viacerých základných ukazovateľov, približne rovnako závažných, ktoré spoločne vytvárajú komplexný obraz o rozsahu tohoto javu metódami **viacrozmernej porovnávacej analýzy**.

V súčastnosti sa na analýzy trhu práce na celoštátnej a okresnej úrovni využívajú vybrané metódy viackriteriálneho porovnávania a metódy zhlukovej analýzy.

Vybrané metódy viackriteriálneho porovnávania[1]

- metóda poradí,
- metóda bodovacia,
- metóda vzdialenosti od fiktívneho objektu.

Cieľom týchto metód je nahradiť niekoľko vybraných ukazovateľov, podľa ktorých chceme porovnať sledované objekty, v našom prípade okresy, jedným syntetickým (agregovaným) ukazovateľom, podľa ktorého potom objekty porovnávame, resp. usporiadame od „najlepšieho“ po „najhorší“. Tým sa viacrozmerný problém hodnotenia mení na problém jednorozmerný.

Ukazovatele, podľa ktorých chceme objekty porovnávať, sú obyčajne veľmi rôznorodé a ich jednoduchý súčet nemá zmysel. Aby sme ich mohli sčítavať (agregovať), musíme ich normovať, teda pretransformovať tak, aby ich súčet pre každý sledovaný objekt mal zmysel.

Najskôr však musíme zaradiť každý sledovaný ukazovateľ do jedného z nasledovných typov.

- **Typ „+“** – žiadúce sú čo najvyššie hodnoty ukazovateľa,
- **Typ „-“** – žiadúce sú čo najnižšie hodnoty ukazovateľa.

Nasledovne popíšem bližšie jednotlivé vybrané metódy:

- **Metóda poradí**

Skutočné hodnoty x_i každej premennej nahradíme ich poradím od 1 po n v usporiadanom rade hodnôt od najnižšej po najvyššiu pri premenných typu „–“ a od najvyššej po najnižšiu pri type „+“. Pri obidvoch typoch premenných potom dosiahneme podľa každého ukazovateľa usporiadanie od „najlepšieho“ (poradie 1) po „najhorší“ objekt (poradie n). Potom z takto vypočítaných poradí podľa každého ukazovateľa vypočítame jednoduchý aritmetický priemer, priemerné poradie podľa všetkých sledovaných ukazovateľov, ktoré bude syntetickou premennou, podľa ktorej nakoniec objekty usporiadame. Najlepší objekt s poradím 1 bude ten, ktorého priemerné poradie je najnižšie, najhorší s poradím n bude ten, ktorého priemerné poradie je najvyššie.

- **Metóda bodovacia**

Podľa tejto metódy pre každý objekt (*i = 1, 2… n*) a pre každú hodnotu ukazovateľa x_{ij} (j = 1, 2 … k) priradíme počet bodov b_{ij} podľa vzťahu:

$$b_{ij} = \frac{x_{ij}}{x_{\max.j}} \times 100,$$ ak je žiadúce aby ukazovateľ dosahoval čo najvyššie hodnoty;

$$b_{ij} = \frac{x_{\min.j}}{x_{ij}} \times 100,$$ ak sú žiadúce čo najnižšie hodnoty ukazovateľa.

Potom pre každý objekt vypočítame súčet bodov za všetky ukazovatele a vydelíme ho počtom ukazovateľov. Tým pre každý objekt dostaneme priemerný počet dosiahnutých bodov. Nakoniec určíme poradie objektov podľa priemerného počtu bodov. Najlepší bude objekt s najvyšším priemerným počtom bodov (poradie 1), najhorší, s poradím n, bude objekt s najnižším priemerným počtom bodov. Počet bodov u jednotlivých ukazovateľov vyjadruje vlastne počet percent, ktorý objektu prislúcha z optimálnej hodnoty ukazovateľa. Priemerný počet bodov vyjadruje, koľko precent prislúcha objektu z maximálne dosažiteľného počtu bodov.

- **Metóda vzdialenosti od fiktívneho objektu**

Fiktívnym objektom rozumieme akýsi abstraktný objekt, ktorý by dosahoval najlepšie hodnoty každého ukazovateľa, teda minimálnu hodnotu pri všetkých ukazovateľoch typu „-“ a maximálnu hodnotu pri ukazovateľoch typu „+“. Ak označíme takúto optimálnu hodnotu j-teho ukazovateľa x_{oj}, potom pre každý objekt vypočítame priemernú vzdialenosť d_i od fiktívneho objektu podľa vzťahu:

$$d_i = \frac{1}{k} \sum_{j=1}^{k} \left| x_{ij} - x_{oj} \right|,$$

pre $i = 1, 2, \ldots n$.

Konečné poradie určíme tak, že najlepší objekt s poradím 1 bude ten, ktorý má najmenšiu vzdialenosť od optimálneho objektu a najhorší s poradím n bude objekt najviac vzdialený od fiktívneho objektu, teda s najvyššou hodnotou d_i.

Pretože jednotlivé ukazovatele môžu byť vyjadrené v rôznych merných jednotkách a aj pri vyjadrení v rovnakých jednotkách môžu mať rôzne ukazovatele odlišnú úroveň hodnôt, je vhodnejšie pracovať s normovanými veličinami, čím sa zaistí lepšia porovnateľnosť všetkých hodnôt x_{ij} podľa vzťahu:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}.$$

Kapitola 2

Metódy zhlukovej analýzy

Cieľ zhlukovej analýzy

Cieľom zhlukovej analýzy je rozčleniť skúmané objekty (v našom prípade okresy) do vnútorne rovnorodých, ale vzájomne výrazne odlišných skupín, tzv. **zhlukov**, na základe zistených hodnôt viacerých premenných. Ide teda o rozklad daného súboru objektov do niekoľkých relatívne homogénnych podsúborov – zhlukov tak, aby si objekty vo vnútri jednotlivých zhlukov boli čo najviac podobné a objekty patriace do rôznych zhlukov si boli podobné čo najmenej. Pritom je každý objekt popísaný skupinou ukazateľov (premenných).

Rovnako ako v mnoho ostatných úlohách z viacrozmernej analýzy máme k dispozícii dátovú maticu $\boldsymbol{X}$ typu $n \times p$, kde n je počet objektov a p je počet premenných. Uvažujeme rôzne rozklady $S^{(k)}$ množiny n objektov do k zhlukov a hľadáme taký rozklad, ktorý by bol z určitého hľadiska najvýhodnejší. Tu pripúšťame len rozklady s disjunktnými zhlukmi. Cieľom je v podstate dosiahnuť stav, keď objekty vo vnútri zhluku sú si podobné čo najviac a s objektami z iných zhlukov čo najmenej. Špecifikácii pojmu podobnosť sa budeme venovať neskôr, teraz si len všimneme, že pojem sa v konkrétnych úlohách vzťahuje len k p premenným, ktoré skúmame. Výber množiny premenných rozhoduje o úspechu analýzy a je nutné mu venovať náležitú pozornosť.

Postupy používané v zhlukovej analýze môžme klasifikovať z viacerých uhlov pohľadu. Najprv rozlíšime:

- úlohy so zadaným počtom zhlukov,
- úlohy, kde určenie vhodného počtu zhlukov je súčasť zadania.

Ak nie je treba obmedzovať objem výpočtov, môžeme v druhom prípade prebrať a vyhodnotiť všetky reálne možnosti voľby k. Ďalej možeme použiť niektorý z algoritmov hierarchického zhlukovania. Ďaľším hľadiskom, podľa ktorého môžeme klasifikovať zhlukovacie algoritmy, je spôsob posudzovania podobnosti. V rôznych etapách algoritmu posudzujeme podobnosť dvoch objektov, podobnosť objektu a zhluku a podobnosť dvoch zhlukov. Rôzne miery podobnosti spomenieme neskôr.

Kritéria pre posúdenie kvality rozkladu

Závažnou všeobecnou otázkou je posúdenie, do akej miery boli v danej situácii pri použití konkrétneho algoritmu, dosiahnuté ciele zhlukovej analýzy. Pre tento účel bolo navrhnutých niekoľko kritérií. V najpoužívanejších kritériách vystupujú nasledovné charakteristiky[4].

- matica vnútrozhlukovej variability:

$$\mathbf{E} = \sum_{h=1}^{k} \sum_{i=1}^{n_h} (\mathbf{x}_{hi} - \bar{\mathbf{x}}_h)(\mathbf{x}_{hi} - \bar{\mathbf{x}}_h)^T;$$

- matica medzizhlukovej variability:

$$\mathbf{B} = \sum_{h=1}^{k} n_h (\bar{\mathbf{x}}_h - \bar{\mathbf{x}})(\bar{\mathbf{x}}_h - \bar{\mathbf{x}})^T;$$

ktoré v súčte dávajú maticu celkovej variability:

$$\mathbf{T} = \sum_{h=1}^{k} \sum_{i=1}^{n_h} (\mathbf{x}_{hi} - \bar{\mathbf{x}})(\mathbf{x}_{hi} - \bar{\mathbf{x}})^T;$$

Základným cieľom je vytvorenie vzájomne vzdialených, ale kompaktných zhlukov. Tento cieľ dosiahneme, ak bude dosiahnuté minimum celkového súčtu štvorcov odchýliek všetkých hodnôt od príslušných zhlukových priemerov.

Vzdialenosť a podobnosť objektov

Podobnosť, resp. nepodobnosť objektov posudzujeme pomocou vybranej funkcie vzdialenosti. Pre výpočet vzdialenosti dvoch objektov podľa p rôznych premenných môžeme použiť viaceré vzťahy. Najčastejšie sa v praxi používa **euklidovská vzdialenosť**. Euklidovskú vzdialenosť medzi i-tym a j-tym objektom na základe hodnôt p rôznych premenných vyjadruje vzorec:

$$D_E(\mathbf{x}_i, \mathbf{x}_{i'}) = \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2}.$$

Pri použití euklidovskej vzdialenosti dochádza niekedy k udalosti že premenná s väčšou varianciou má väčší vplyv na zhlukovanie (ak máme premenné s rôznymi rozptylmi). Doporučuje sa štandardizácia dát pred výpočtom euklidovskej vzdialenosti, čím sa vlastne dosiahne tzv. Mahalanobisova vzdialenosť daná vzorcom[4]:

$$D_E^2(\mathbf{u}_i,\mathbf{u}_{i'}) = D_M^2(\mathbf{x}_i,\mathbf{x}_{i'}) = (\mathbf{x}_i - \mathbf{x}_{i'})'\mathbf{S}^{-1}(\mathbf{x}_i - \mathbf{x}_{i'})$$

Musíme však zvážiť, že pri jej použití potlačíme vplyv rozdielov vo variabilite premenných na výsledky, čo nemusí byť žiadúce.

Vzdialenosť medzi zhlukmi

V prvom kroku vypočítame maticu vzdialeností medzi všetkými dvojicami objektov. V prípade n objektov bude matica typu n x n.V ďalsích krokoch sa zhlukovanie riadi niektorou metódou zhlukovania, založenou na mierach vzdialenosti medzi zhlukmi.

- **Metóda najbližšieho suseda**

Kritériom pre spojovanie zhlukov je minimum z q možných medzizhlukových vzdialeností objektov. Pri úprave matice vzdialeností pre ďalší cyklus použijeme nasledovný vzorec: $D_{gg'} = \min(D_{g'h}, D_{g'h'})$.

Pri použití tejto metódy sa často i dosť vzdialené objekty môžu spojiť v rovnakom zhluku, pokiaľ väčší počet ďaľších objektov medzi nimi vytvorí akýsi most. Toto charakteristické reťazenie objektov sa považuje za nevýhodu, najmä keď máme dôvod požadovať, aby zhluky mali obvyklý eliptický tvar so zhusteným jadrom.

- **Metóda najvzdialenejšieho suseda**

Ako vyplýva z názvu, kritériom pre spojovanie zhlukov je maximum z q možných medzizhlukových vzdialeností objektov. Pri úprave matice vzdialeností postupujeme podľa vzorca: $D_{gg'} = \max(D_{g'h}, D_{g'h'})$.

Nežiadúci reťazový efekt v tomto prípade odpadá, naopak je tu tendencia k tvorbe kompaktných zhlukov, avšak nijako mimoriadne veľkých.

- **Metóda priemernej väzby**

Táto metóda používa ako kritérium pre spojovanie zhlukov priemer z q možných medzizhlukových vzdialeností objektov. Pri prepočítavaní matice vzdialeností použijeme vzorec:

$$D_{gg'} = \frac{n_h D_{g'h} + n_{h'} D_{g'h'}}{n_h + n_{h'}}.$$

Metóda vedie často k podobným výsledkom ako metóda najvzdialenejšieho suseda.

- **Centroidná metóda**

V tejto metóde už nevychádzame zo zhrňovania informácií o medzizhlukových vzdialenostich objektov. Kritérium je euklidovská vzdialenosť centroidov

$$D_E(\bar{x}_h, \bar{x}_{h'}) = \sum_{j=1}^{p} (\bar{x}_{hj} - \bar{x}_{h'j})^2$$

Prepočítanie matice vzdialeností sa prevedie podľa vzťahu:

$$D_{gg'} = \frac{1}{n_h + n_{h'}} \left(n_h D_{g'h} + n_{h'} D_{g'h'} - \frac{n_h n_{h'}}{n_h + n_{h'}} D_{hh'} \right).$$

- **Wardova metóda**

Najprv sa spoja do jedného zhluku také dva objekty, ktorých vzdialenosť je najmenšia. Pri ďalšom postupe spojíme dva zhluky alebo samostatné objekty, ktoré považujeme za jednoprvkové zhluky, ktoré majú najmenšiu medzizhlukovú vzdialenosť. Postup zhlukovania končí po (n – 1) krokoch tým, že všetky objekty budú patriť do jedného zhluku.

Algoritmus pre vytvorenie hierarchickej postupnosti rozkladov[4]

K najpoužívanejším postupom uplatňovaným v zhlukovej analýze patrí vytváranie hierarchickej postupnosti rozkladov tak, že:

I. Vypočítame maticu **D** vhodných mier vzdialeností.

II. Začneme proces od rozkladu $\mathbf{S}^{(n)}$, tj. od *n* zhlukov, z katorých každý obsahuje jeden objekt.

III. Prehľadáme maticu **D** a nájdeme dva zhluky (*h*-ty a *h´*-ty), ktorých vzdialenosť je minimálna (v prípade metódy najbližšieho suseda).

IV. Spojíme *h*-ty a *h´*-ty zhluk do nového *g*-teho zhluku. V matici **D** vymažeme *h*-ty a *h´*-ty riadok aj stĺpec a nahradíme ho riadkom a stĺpcom pre nový zhluk.

V. Poznačíme si poradie cyklu l=1, 2,…, *n*-1, identifikáciu spojených objektov *h*, *h´*a hladinu pre spojenie $d_l = D_{hh'}$.

VI. Pokiaľ proces vytvárania rozkladu už neskončil spojením všetkých objektov do jediného zhluku $\mathbf{S}^{(1)}$, pokračujeme krokom III.

Popísali sme aglomeratívny hierarchický postup. Menej používaný je opačný, divizívny hierarchický postup, pri ktorom vychádzame z jediného zhluku $\mathbf{S}^{(1)}$ a v každom kroku jeden zo zhlukov rozštiepime na dva, takže na konci procesu dostávame $\mathbf{S}^{(n)}$.

Názornou pomôckou pri vyhodnocovaní výsledkov zhlukovej analýzy je grafické znázornenie postupu zhlukovania tzv. **dendrogram**. V ňom sa na vodorovnú os zobrazujú objekty tak, ako sa postupne zhlukovali a na zvislú os hodnoty vzdialenosti, pri ktorých sa zhlukovali. Z dendrogramu vidíme, v ktorých skupinách okresov je veľmi, v ktorých menej a v ktorých veľmi málo podobná situácia podľa zvolených ukazovateľov. Vyjadruje to úroveň, na ktorej sa okresy zhlukovali. Čím je úroveň zhlukovania vyššia, tým je menšia podobnosť okresov v danom zhluku.

Nakoniec treba ešte spomenúť, že rôzne zhlukovacie algoritmy a rôzne vzdialenosti dávajú vo všeobecnosti rôzne výsledky. Program S+, pomocou ktorého som spracoval analýzy, ponúka v zhlukovej analýze aglomeratívny hierarchický postup s rôznymi voľbami vzdialeností.

Metódy faktorovej analýzy

Cieľ faktorovej analýzy

Všeobecným cieľom metód faktorovej analýzy je nájsť spôsob, ktorý by umožňoval sumarizáciu väčšieho počtu originálnych premenných do menšej skupiny skrytých faktorov tak, aby strata informácií bola čo najmenšia.

Faktorová analýza skúma vzájomné vzťahy medzi ukazovateľmi a snaží sa ich vysvetliť ako dôsledok pôsobenia skrytých spoločných dimenzií (faktorov). Každý ukazovateľ je považovaný za závisle premennú, ktorá je funkciou určitej skupiny skrytých (latentných) faktorov. Faktorová analýza plní tieto funkcie:

- identifikuje skupinu dimenzií, ktoré nie sú vo veľkej skupine pôvodných premenných priamo pozorovateľné,
- identifikuje užšiu skupinu pôvodných premenných, ktorá môže byť následne použitá v zhlukovej alebo regresnej analýze,
- vytvára menší súbor úplne nových premenných, ktoré môžu čiastočne alebo úplne nahradiť pôvodné premenné v následnej zhlukovej alebo regresnej analýze.

Aplikácia faktorovej analýzy prebieha v určitých samostatných krokoch. Každý krok sa vyznačuje riešením a zodpovedaním špecifických otázok. Najprv uvedieme model faktorovej analýzy.

Model faktorovej analýzy

Nech **x** je *p*-rozmerný vektor pozorovaných premenných s kovariančnou maticou Σ. Môžeme predpokladať, že jednotlivé premenné sú centrované, to znamená, že stredné hodnoty sú nulové. Vo faktorovej analýze sa vysvetľuje vzájomná lineárna závislosť pozorovaných premenných X_j, j = 1, 2, …, p, existenciou menšieho počtu nepozorovateľných veličín f_1, f_2, …, f_m (nazývaných spoločné faktory) a *p* ďaľších zdrojov variability ε_1, ε_2, …, ε_p (nazývaných chybové či špecifické faktory alebo tiež rušivé či reziduálne zložky). Spoločné faktory vyvolávajú korelácie medzi premennými, zatiaľ čo chybové faktory prispievajú k rozptylu jednotlivých pozorovaných premenných. Predmetom záujmu faktorovej analýzy sú predovšetkým spoločné faktory. Základný predpoklad faktorovej analýzy môžeme zapísať ako:

$$
\begin{aligned}
X_1 &= \lambda_{11} f_1 + \lambda_{12} f_2 + \ldots + \lambda_{1m} f_m + \varepsilon_1 \\
X_2 &= \lambda_{21} f_1 + \lambda_{22} f_2 + \ldots + \lambda_{2m} f_m + \varepsilon_2 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
X_p &= \lambda_{p1} f_1 + \lambda_{p2} f_2 + \ldots + \lambda_{pm} f_m + \varepsilon_p
\end{aligned}
$$

alebo v maticovom zápise ako $\mathbf{x} = \Lambda \mathbf{f} + \varepsilon$. Koeficienty λ_{jk} sa nazývajú faktorové váhy. O matici faktorových váh Λ sa predpokladá, že má hodnosť m.

Ďaľšie predpoklady faktorovej analýzy sa týkajú nepozorovateľných náhodných vektorov **f** a ε. O spoločných faktoroch sa predpokladá, že sú to normované veličiny s nulovými strednými hodnotami a s jednotkovými rozptylmi. V ortogonálnom faktorovom modeli sa navyše predpokladá, že spoločné faktory sú nekorelované. O prvkoch vektora ε, teda o chybových faktoroch sa predpokladá, že majú nulové stredné hodnoty, sú navzájom nekorelované a navyše sú nezávislé na spoločných faktoroch. Ak označíme Ψ_j ako rozptyl *j*-teho chybového faktoru, môžeme uvedené predpoklady stručne zapísať:

1. $E(\varepsilon) = \mathbf{o}_p$,
2. $E(\varepsilon\varepsilon') = \Psi$, kde Ψ je diagonálna matica s prvkami Ψ_j na diagonále,
3. $E(\boldsymbol{f}) = \mathbf{o}_m$,

4. $E(\boldsymbol{f}\boldsymbol{f}') = \mathbf{I}_m$,

5. $E(\boldsymbol{f}\varepsilon') = \mathbf{O}_{m,\,p}$. (2)

Z predpokladu (1) a (2) ortogonálneho faktorového modelu vyplýva, že kovariančnú maticu Σ vektorov pozorovaných premenných môžeme písať ako:

$$\Sigma = E(\mathbf{x}\mathbf{x}') = E[(\Lambda\boldsymbol{f} + \varepsilon)(\Lambda\boldsymbol{f} + \varepsilon)'] = \Lambda E(\boldsymbol{f}\boldsymbol{f}')\Lambda' + E(\varepsilon\varepsilon') = \Lambda\Lambda' + \Psi. \qquad (3)$$

Za predpokladu platnosti faktorového modelu môžeme kovariančnú maticu pozorovaných premenných rozložiť na dve časti, z ktorých prvá je kovariančná matica vektorov $\Lambda\mathbf{f}$ a druhá je kovariančná matica chybových faktorov. Ako vyplýva z (3), rozptyl j-tej premennej X_j môžeme písať ako:

$$D(X_j) = \sigma_j^2 = \sum_{k=1}^{m} \lambda_{jk}^2 + \Psi_j,$$

a kovarianciu j - tej a j' - tej premennej, $j \neq j'$, ako:

$$C(X_j, X_{j'}) = \sigma_{jj'} = \sum_{k=1}^{m} \lambda_{jk}\lambda_{j'k}.$$

Matica $\Lambda\Lambda'$ sa nazýva redukovaná kovariančná matica a jej diagonálne prvky:

$$\sum_{k=1}^{m} \lambda_{jk}^2 = \sigma_j^2 - \Psi_j = h_j^2$$

sa nazývajú komunality premenných. Komulita premennej X_j udáva tú časť rozptylu, ktorá je vysvetlená pôsobením spoločných faktorov. Ostávajúca časť rozptylu X_j, tj. Ψ_j, je nazývaná špecifickým (chybovým) rozptylom premennej X_j.

Z faktorového modelu ďalej vyplýva, že parameter λ_{jk} predstavuje kovarianciu *j*-tej premennej s *k*-tym spoločným faktorom. Túto skutočnosť ľahko overíme, ak zapíšeme kovariančnú maticu pozorovaných a faktorových veličín ako:

$$E(\mathbf{x}\boldsymbol{f}') = E((\Lambda\boldsymbol{f} + \varepsilon)\boldsymbol{f}') = \Lambda E(\boldsymbol{f}\boldsymbol{f}') + E(\varepsilon\boldsymbol{f}') = \Lambda$$

Ak je **R** korelačná matica, potom hodnoty λ_{jk} predstavujú korelačné koeficienty premenných so spoločnými faktormi.

Príprava a posúdenie vhodnosti dát

- **definovanie výberového súboru**

Faktorová analýza je vhodnou metódou, ak cieľom výskumu je objasnenie príčin korelácie medzi použitými premennými a súčasne redukcia ich počtu. V prvom kroku jej realizácie je potrebné určiť:

- *ktoré premenné budú predmetom analýzy,*
- *počet analyzovaných premenných,*
- *typy a merné jednotky analyzovaných premenných,*
- *veľkosť výberového súboru.*

Všeobecne najvhodnejšie pre faktorovú analýzu sú kvantitatívne premenné, v špecifických prípadoch možno použiť aj poradové alebo umelé premenné. Pre účely faktorovej analýzy sa odporúča, aby veľkosť súboru bola 100 pozorovaní a viac. V ideálnych podmienkach možno požadovať pri veľkosti vzorky dodržanie konzervatívneho pravidla. Jeho splnenie predpokladá, aby počet pozorovaní bol štyri až päťkrát väčší ako počet analyzovaných premenných. Pri reálnych ekonomických analýzach sa možno uspokojiť aj s pomerom 2 : 1 medzi počtom pozorovaní a počtom premenných. Pre menšom rozsahu výberového súboru možno získané výsledky interpretovať len s veľkou opatrnosťou.

- **analýza korelačnej matice**

Faktorová analýza sa najčastejšie používa na hodnotenie vzťahov medzi premennými. Možno ju využiť aj na klasifikáciu objektov analýzy. Východiskom pre ďalšie fázy je korelačná matica premenných. Na jej základe možno získať prvú predstavu o vzťahoch medzi premennými a posúdiť celkovú vhodnosť dát pre faktorovú analýzu. Väčšina postupov ktoré sa pre tento účel používajú, je zásadne heuristická, ponechávajúca konečné rozhodnutie na úsudok a skúsenosti analytika.

Vysoká korelácia medzi premennými poukazuje na to, že premenné môžu byť rozdelené do homogénnych, relatívne nezávislých skupín. Možno tiež predpokladať, že príčinou takejto štruktúry je existencia skrytých faktorov alebo dimenzií. Naopak nízke koeficienty korelácie naznačujú, že premenné nemajú veľa spoločného a použitie faktorovej analýzy je nevhodné.

Exaktnejšiu odpoveď na rovnakú otázku poskytuje výpočet Bartletovho testu sférickosti. Testuje nulovú hypotézu, že korelačná matica je jednotkovou maticou, čo je analogické s predpokladom, že premenné sú navzájom nezávislé. Ak túto nulovú hypotézu nemožno zamietnuť, faktorová analýza nie je vhodná. Nedostatkom testu je vysoká citlivosť na veľkosť výberového súboru. Pri veľkom rozsahu výberového súboru vedie k zamietnutiu formulovanej hypotézy, aj keď koeficienty korelácie sú veľmi malé. Test požaduje, aby výberové dáta pochádzali z viacrozmerného normálneho rozdelenia. V programe S+ sa uvedený test nenachádza.

Pre rovnaké účely možno vypočítať Kaiser–Meyer–Olkinovu mieru adekvátnosti výberových dát. Tiež tzv. KMO-index je vlastne indexom porovnávajúcim veľkosť zistených koeficientov korelácie a veľkosť parciálnych koeficientov korelácie. Autori[2] tejto miery definovali len intuitívne hranice pre vhodnosť dát. Program S+ túto hodnotu neponúka.

Hodnota KMO-indexu	Hodnotenie vhodnosti dát
≥0,90	Dokonalé
0,80+	Pozoruhodné
0,70+	Náležité
0,60+	Obyčajné
0,50+	Slabé
≤0,50	Nevhodné

Odhad modelu faktorovej analýzy pred rotáciou

- **výber metódy odhadu faktorov**

Pre odhad modelu faktorovej anayzy možno použiť nasledovné metódy:

- *metódu hlavných komponentov*
- *metódu hlavných faktorov*
- *kanonickú faktorovú analýzu (metóda maximálnej vierohodnosti)*
- *alfa faktorovú analýzu*
- *image faktorovú analýzu*

Najpoužívanejšie sú metóda hlavných komponentov, metóda maximálnej vierohodnosti (MLE) a metóda hlavných faktorov, ktorá sa tiež nazýva modifikovaná metóda hlavných komponentov. V tomto kroku sa teda problém zužuje na rozhodnutie, či model bude odhadnutý ako model analýzy hlavných komponentov alebo ako model analýzy hlavných faktorov. Výber je závislý od cieľa analýzy. Model hlavných komponentov je vhodný pre sumarizáciu maximálnej pôvodnej informácie do minimálneho počtu faktorov s cieľom ďaľšieho prognostického použitia. Metóda analýzy hlavných faktorov sa používa na identifikáciu skrytých nepozorovateľných dimenzií. Voľba vhodnej metódy si vyžaduje poznatky o typoch variability. Celkový rozptyl každej premennej predstavuje diagonálny prvok korelačnej (kovariančnej) matice v príslušnom riadku. Pre účely faktorovej analýzy sa táto celková variabilita rozkladá na dve zložky:

(a) komunalitu, ktorá je vysvetlená pôsobením spoločných faktorov,

(b) unicitu, ktorá je spôsobená špecifickým faktorom a pre každú premennú je jedinečná.

Obidve spomenuté metódy pri odhade faktorov vychádzajú z tzv. redukovanej korelačnej matice, ktorá je odvodená z korelačnej matice premenných.V nej sa diagonálne prvky nahradia odhadmi komunalít jednotlivých premenných. Základný koncepčný rozdiel medzi metódami je v potrebe odhadu komunalít.

Metóda hlavných komponentov predpokladá, že počiatočné komunality všetkých premenných sú rovné jednej. Hypotéza, že komunality sú rovné jednej, vlastne zároveň naznačuje, že v modeli sa neuvažuje prítomnosť špecifických faktorov.Počet odhadnutých komponentov je rovnaký ako počet premenných Predpokladá sa, že malý počet hlavných komponentov vysvetlí veľkú časť variability v dátach. Tieto komponenty sa potom považujú za spoločné skryté faktory. Variabilita každej premennej, ktorá je vysvetlená vplyvom spoločných komponentov, sa prezentuje ako komunalita. Zvyšná časť variability jednotlivých premenných je unicita.

Metóda hlavných faktorov spočíva v aplikácii metódy hlavných komponentov na redukovanú kovariančnú (korelačnú) maticu premenných. Ak je faktorový model $\Sigma=\Lambda\Lambda' + \Psi$ správne špecifikovaný, malo by m spoločných faktorov vysvetlovať mimodiagonálne prvky matice Σ a ďalej časti diagonálnych prvkov, odpovedajúcich komunalitám jednotlivých premenných. Platí $\sigma_j^2 = h_j^2 + \Psi_j$, kde h_j^2 predstavuje komunalitu j-tej premennej X_j. Ak odčítame od diagonálnych prvkov matice Σ príspevky špecifických faktorov k rozptylu premenných, má výsledná redukovaná kovariančná matica tvar $\Sigma - \Psi = \Lambda\Lambda'$.

Predpokladajme, že disponujeme určitými počiatočnými odhadmi špecifických rozptylov Ψ_j^*. Ak nahradíme j–ty diagonálny prvok matice **S** (výberová kovariančná matica, ktorá je odhadom neznámej matice Σ) hodnotou $h_j^{*2} = s_j^2 - \Psi_j^*$, dostávame redukovanú výberovú kovariančnú maticu:

$$\mathbf{S}_r = \begin{bmatrix} h_1^{*2} & s_{12} & \dots & s_{1p} \\ s_{21} & h_2^{*2} & \dots & s_{2p} \\ \dots & \dots & \dots & \dots \\ s_{p1} & s_{p2} & \dots & h_p^{*2} \end{bmatrix}$$

Ak prehliadneme nepresnosti spôsobené výberom, potom m spoločných faktorov by malo vysvetlovať všetky prvky redukovanej výberovej kovariančnej matice $\mathbf{S}_r$. Maticu $\mathbf{S}_r$ môžeme rozložiť ako

$$\mathbf{S}_r = \mathbf{L}_r^* \mathbf{L}_r^{*\prime},$$

Kde $\mathbf{L}_r^*$ (obsahuje prvky l_{jk}^*) je matica počiatočných odhadov faktorových záťaží.

Metóda hlavných komponentov berie za odhad parametrov faktorového modelu hodnoty

$$\mathbf{L}_r^* = \left\lfloor \sqrt{l_{(1)}^*}\mathbf{v}_1^* \sqrt{l_{(2)}^*}\mathbf{v}_2^* \dots \sqrt{l_{(3)}^*}\mathbf{v}_3^* \right\rfloor$$

$$\Psi_j^* = 1 - \sum_{k=1}^{m} l_{jk}^{*2},$$

kde $l^*_{(k)}$, k = 1, 2, …, m, predstavuje m najväčších charakteristických čísel redukovanej výberovej kovariančnej matice $\mathbf{S}_r$ a $\mathbf{v}_k$ sú k nim príslušné charakteristické vektory.

Na základe odhadu Ψ_j^* môžeme urobiť nový odhad komunalít ako

$$h_j^{*2} = \sum_{k=1}^{m} l_{jk}^{*2}.$$

Metóda hlavných faktorov postupuje ďalej iteratívne. Dosadením odhadu komunalít na diagonálu matice **S** sa spresní redukovaná kovariančná matica a pristúpi sa k novému výpočtu charakteristických čísel. Iteratívny postup konverguje po niekoľkých krokoch.

Metóda maximálnej vierohodnosti vyžaduje predpoklad že $\boldsymbol{f}$ a ε majú viacrozmerné normálne rozdelenie s nulovými vektormi stredných hodnôt a kovariančnými maticami $\mathbf{I}_m$ a Ψ. Odhad metódou maximálnej vierohodnosti vychádza z prvkov matice **S**. Za predpokladu *m*-faktorového modelu má logaritmus vierohodnostnej funkcie prvkov matice Λ a Ψ tvar

$$\ln L(\Lambda,\Psi) = -\tfrac{1}{2}(n-1)[\ln\left|\Lambda\Lambda' + \Psi\right| + st(\Lambda\Lambda' + \Psi)^{-1} S$$

Pre získanie maximálne vierohodných odhadov parametrov faktorového modelu je treba maximalizovať logaritmus vierohodnostnej funkcie vzhľadom k *p(m+1)* prvkom matíc Λ a Ψ. V súčastnosti existujú efektívne výpočtové postupy, ktoré umožňujú pomerne ľahko získať maximálne vierohodné odhady parametrov faktorového modelu.

Počiatočný tvar riešenia získaného metódou maximálnej vierohodnosti odpovedá podmienke jednoznačnosti tejto metódy, ktorá požaduje, aby matica $\Lambda'\Psi^{-1}\Lambda$ bola diagonálna.

- **odhad počiatočných komunalít**

V prípade že explicitne možno predpokladať existenciu faktorov, je potrebné odhadnúť počiatočnú mieru, s akou ovplyvňujú konkrétnu premennú. Kvalita odhadu počiatočných komunalít ovplyvňuje celkovú kvalitu odhadnutého modelu faktorovej analýzy. Diagonálne prvky redukovanej matice možno odhadnúť:

- *iteračným postupom* (opakovanou extrakciou faktorov),
- *najvyšším koeficientom korelácie danej premennej s ostatnými,*
- *priemerným koeficientom korelácie* (stĺpca/riadku korelačnej matice),
- *štvorcom viacnásobného koeficienta determinácie premenných,*
- *najvyššou koreláciou* (ktorá sa vypočíta ako pomer druhej mocniny súčtu konkrétneho stĺpca korelačnej matice ku celkovému súčtu všetkých koeficientov korelácie).

Pre dostatočne veľký rozsah analyzovaného súboru sú rozdiely vo výsledkoch pri rôznych odhadoch počiatočných komunalít minimálne.

- **určenie počtu faktorov**

Faktorová analýza umožňuje odhadnúť maximálne toľko faktorov, koľko je premenných. Cieľom však je vysvetliť maximum korelácie medzi premennými pomocou minimálneho počtu faktorov. Vzniká problém, koľko faktorov z odhadnutého maximálneho počtu postačuje na vysvetlenie korelácie medzi premennými. Ak nevyberieme všetky faktory, zákonite dochádza ku strate informácií obsiahnutých v dátach. Musíme sa rozhodnúť, akú stratu informácií sme ochotní akceptovať. Podporne možno použiť nasledovné kritériá[3]:

- *apriórna požiadavka na počet faktorov*

  V niektorých prípadoch môže byť dopredu známy počet skrytých faktorov, napríklad z predchádzajúcich analýz.
- *apriórna požiadavka na percento vysvetlenej variability*

  Počet štatisticky významných faktorov závisí od toho, koľko percent z celkovej variability vystihujú.
- *posúdenie vlastných čísel*

  Pre ďalšiu analýzu je vhodné ponechať len faktory, ktorých vlastné čísla (časť celkovej variability vysvetlená daným faktorom) sú väčšie ako jedna.
- *scree plot*

  Je to graf zobrazujúci poradové čísla faktorov a vlastné čísla faktorov. Spojnica týchto bodov má klesajúcu tendenciu. Sledujeme jeho priebeh a hľadáme v ňom výraznejší zlom. Akceptuje faktory, ktoré sa v grafe nachádzajú pred týmto zlomom.

Metóda maximálne vierohodného odhadu parametrov faktorového modelu umožňuje testovať hypotézu, že m-faktorový model je vhodný pre vysvetlenie napozorovaných kovariancií alebo korelácií. Testovaná hypotéza má tvar:

$$\mathbf{H} : \Sigma = \Lambda\Lambda' + \Psi,$$

kde Λ je typu p x m.

Použitie vierohodnostného pomeru vedie k testovaciemu kritériu:

$$V_1 = \left[n - 1 - \tfrac{1}{6}(2p+5) - \tfrac{2}{3}m\right] \ln \frac{\left|\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}\right|}{|\mathbf{S}|},$$

kde označenia so strieškou sú maximálne vierohodné odhady Λ a Ψ; **S** je výberová kovariančná matica. Pokiaľ platí hypotéza **H**, má veličina V_1 (pre veľké n) rozdelenie χ^2 s $[(p - m)^2 - p - m]/2$ stupňami voľnosti. Z vlastnosti invariancie plynie, že testovacie kritérium môžeme použiť aj v prípade, keď bola faktorová analýza urobená na základe korelačnej matice. Vzhľadom k tomu, že počet stupňov voľnosti veličiny V_1 musí byť kladný, počet spoločných faktorov musí vyhovovať podmienke:

$$m < \tfrac{1}{2}(2p + 1 - \sqrt{8p + 1}).$$

Všimnime si interpretáciu veličiny V_1. Logaritmus pomeru determinantov v definícii veličiny môžeme aproximovať výrazom:

$$\sum_{j}^{p}\sum_{j'}^{p}\frac{(s_{jj'} - \hat{\sigma}_{jj'})^2}{\hat{\psi}_j\hat{\psi}_{j'}},$$

kde $\hat{\sigma}_{jj'} = \sum_{k=1}^{m}\lambda_{jk}\lambda_{j'k}, j \neq j'$, predstavuje kovarianciu premenných X_j a $X_{j'}$ reprodukovanú faktorovým modelom. Uvedené kritérium je mierou toho, nakoľko sa reziduálne koeficienty korelácie po získaní *m* spoločných faktorov líšia od nuly.

Popísaný test nemožno využívať mechanicky, pretože pre veľké *n* a relatívne malé *m* (vzhľadom k *p*) môžeme očakávať, že test zamietne hypotézu **H** a príde k zvýšeniu počtu spoločných faktorov.

Rotácia faktorov

Faktorová rotácia umožňuje zoskupenie premenných do nových dimenzií, faktorov. Pôvodné premenné sú do nových faktorov spájané pomocou numerických váh. Tieto váhy sa nazývajú faktorové saturácie. Sú kľúčom k pochopeniu a interpretácii jednotlivých faktorov. Druhá mocnina faktorovej saturácie vyjadruje, aká časť z celkovej variability danej premennej je vysvetlená pôsobením konkrétneho faktora. Matica, ktorá obsahuje faktorové saturácie všetkých premenných (riadky matice) so všetkými faktormi (stĺpce matice) sa nazýva matica faktorovej štruktúry.

Matica faktorovej štruktúry získaná konkrétnou metódou odhadu faktorov je iba prvotným nástrojom pre redukciu dát, čiže výber významných faktorov. Výberom štatisticky významných faktorov matica faktorovej štruktúry obsahuje tzv. Nerotované riešenie faktorovej analýzy. To znamená, že po splnení vopred stanoveného kritéria obsahuje najlepšie riešenie v zmysle zahrnutia maximálnej informácie z pôvodných premenných. Vzniká však otázka, či je takéto riešenie nejlepšie aj z pohľadu ekonomickej interpretácie výsledkov a z pohľadu konečného užívateľa.

Ak nemožno nerotované riešenie jednoznačne považovať za vyhovujúce, je vhodné použiť metódy rotácie faktorov na jeho zjednodušenie. Spoločným cieľom všetkých metód rotácie je dosiahnuť lepšiu interpretáciu riešenia pomocou zjednodušenia riadkov a stĺpcov matice faktorovej štruktúry. Stĺpce matice faktorovej štruktúry predstavujú jednotlivé faktory. Riadky obsahujú saturácie jednotlivých premenných. Zjednodušením riadkov sa rozumie dosiahnutie stavu, keď každý riadok obsahuje čo najviac faktorových saturácií blízkych nule. Podobne pri zjednodušení stĺpcov sa snažíme dosiahnuť v každom stĺpci čo najviac faktorových saturácií blízkych nule.

Pri rotácii dochádza k otáčaniu osí jednotlivých faktorov. Použitím rotácie nie sú dotknuté odhady komunalít jednotlivých premenných z nerotovaného riašenia. Existujú nasledovné metódy rotácie:

- *ortogonálna rotácia* je najjednoduchším typom rotácie. Osi faktorov pri nej zvierajú 90 – stupňový uhol. To zároveň znamená, že faktory sú navzájom nezávislé. Všeobecne platí že ortogonálne rotácie sú matematicky jednoduchšie. Ortogonálna rotácia je vhodná pre redukciu veľkého počtu premenných, keď sa nekladie dôraz na zmysluplnosť získaných faktorov a uvažuje sa s ich použitím následne v regresii alebo inej predikčnej metóde. Postupy ortogonálnej rotácie sú prepracovanejšie a pri ďalšej analýze faktorov umožňujú odstrániť multikolinearitu v dátach. Najpoužívanejšie typy ortogonálnej rotácie sú:
    - *VARIMAX* - je to postup zjednodušujúci stĺpce matice faktorovej štruktúry. Každá premenná potom výrazne identifikuje len jeden faktor. Znamená to, že takáto premenná by mala mať s daným faktorom vysokú faktorovú saturáciu a pri ostatných faktoroch by mala mať saturácie blízke nule. Dosiahne sa tým faktorový model, v ktorom každý faktor reprezentuje rozdielnu skutočnosť.
    - *QUARTIMAX* - v porovnaní s typom VARIMAX zjednodušuje riadky matice faktorovej štruktúry. Jej úlohou je identifikovať model, v ktorom všetky premenné pomerne silne identifikujú jeden spoločný faktor a každá z nich samostatne jeden špecifický faktor. Metódu QUARTIMAX je vhodné použiť, ak na základe informácií o analyzovanom jave možno očakávať prítomnosť spoločného faktora.

- *kosouhlá rotácia* - faktorové osi možno otáčať aj tak, že uhol medzi nimi nie je 90 – stupňov. V takom prípade ide o kosouhlú rotáciu. Odhadnuté faktory sú navzájom korelované. Kosouhlé riešenia sú flexibilnejšie (lebo nepredpokladajú ortogonalitu faktorov) a realistickejšie. Ak pôvodné premenné boli významne korelované, je vhodné predpokladať, že nové faktory nebudú nezávislé. Kosouhlá rotácia je vhodná pre získanie zmysluplných realistických faktorov. Tento typ rotácie popisuje zoskupovanie sa premenných presnejšie.

V praxi je nerotované riešenie väčšinou nepostačujúce. Väčšinou sa totiž rotovaním získa riešenie, ktoré je jednoduchšie na účely interpretácie. Účelom rotácie je nájsť zmysluplné faktory a čo najjednoduchšiu faktorovú štruktúru. Štatistický balík S+ ponúka oba typy rotácií. Neexistuje však jasné pravidlo na výber rotácie. Čo sa týka výberu medzi ortogonálnou a kosouhlou, tak v regresii môžme rozhodovať medzi dvoma prípadmi:

1. ak máme veľké množstvo prípadne korelovaných regresorov a ak regresiu robíme za účelom predikcií, tak volíme ortogonálnu,
2. ale ak cieľom je nájsť niekoľko zmysluplných faktorov, tak sa odporúča kosouhlé riešenie.

- **interpretácia faktorov**

Interpretácia odhadnutých faktorov a celého modelu vychádza z faktorových saturácií jednotlivých faktorov po rotácii a znalosti ich vzťahov. Hodnotovo sú faktorové saturácie z intervalu $\langle -1,+1 \rangle$. Vysoká saturácia znamená, že faktor a daná premenná majú veľa spoločného. Na jednej strane daná premenná dobre identifikuje skrytý faktor a na druhej strane významne ovplyvňuje variabilitu danej premennej. Neexistuje však nijaká objektívna hranica, podľa ktorej by bolo možné koeficient ohodnotiť ako vysoký alebo nízky. Možno len akceptovať nasledovné kritériá pre:

- *absulútnu hodnotu faktorových saturácií*

Za významné sa považujú hodnoty vyššie ako 0,30. Saturácie vyššie ako 0,40 sa považujú za stredne významné a vyššie ako 0,50 sa považujú za veľmi významné. Čím vyššia absolútna hodnota faktorovej saturácie, tým významnejšia je premenná pre interpretáciu faktora. Uvedené konkrétne hodnoty má zmysel použiť, ak analyzovaný súbor obsahuje viac ako 50 pozorovaní. Toto kritérium má však nevýhody, neberie do úvahy počet pozorovaní a ani poradie faktora.

- *faktorové saturácie ako koeficienty korelácie*

Faktorové saturácie zo štandardizovaných dát sú koeficientami korelácie faktorov s indikačnými premennými. Ich významnosť možno tiež interpretovať analogicky. Pre hladinu významnosti 5% podľa veľkosti súboru sú štatisticky významné faktorové saturácie uvedené v tabuľke[3]:

Factor loading	Sample size needed for significance
350	0.3
250	0.35
200	0.4
150	0.45
120	0.5
100	0.55
85	0.6
70	0.65
60	0.7
50	0.75

Na základe významných faktorových saturácií možno pristúpiť k pomenovaniu jednotlivých faktorov. Do úvahy treba brať hlavne premenné s veľmi významným vplyvom na faktor. Tieto premenné možno zároveň považovať za najreprezentatívnejšie. Premenné s najvyššími saturáciami najlepšie odrážajú skrytú štruktúru v dátach a môžu byť samostatne použité v ďalšej analýze ako reprezentant konkrétneho faktora.

Analogicky možno v ďalšej analýze použiť faktorové skóre. Faktorové skóre sú hodnoty nových premenných, ktoré boli faktorovou analýzou získané.

V kosouhlej rotácii matica faktorových saturácií už nie je korelačnou (kovariančnou) maticou medzi pôvodnými premennými a faktormi. Za účelmi interpretácie faktorov sa u kosoúhlej rotácii používa matica nákladov. Program S+ vo výstupe kosouhlej rotácie udáva aj túto maticu.

Kapitola 3

Výsledky a interpretácia analýz

VIACKRITERIÁLNE POROVNÁVANIE

Metóda poradí

MESTO	PORADIE
Bratislava II	1
Bratislava V	2
Bratislava IV	3
Bratislava III	4
Pezinok	5
Bratislava I	6
Liptovský Mikuláš	7
Trnava	8
Nové Mesto nad Váhom	9
Banská Bystrica	10
Trenčín	11
Hlohovec	12
Piešťany	13
Púchov	14
Malacky	15
Partizánske	16
Žilina	17
Senec	18
Zvolen	19
Martin	20
Ilava	21
Prievidza	22
Myjava	23
Detva	24
Považská Bystrica	25
Senica	26
Galanta	27
Topoľčany	28
Poprad	29
Žiar nad Hronom	30
Košice IV	31
Dolný Kubín	32
Tvrdošín	33
Ružomberok	34
Košice II	35
Nitra	36
Banská Štiavnica	37
Bytča	38
Poltár	39
Čadca	40
Košice I	41
Skalica	42
Bánovce nad Bebravou	43
Svidník	44
Košice III	45
Turčianske Teplice	46
Vranov nad Topľou	47
Žarnovica	48
Bardejov	49
Prešov	50
Dunajská Streda	51
Námestovo	52
Nové Zámky	53
Komárno	54
Snina	55
Krupina	56
Zlaté Moravce	57
Levice	58
Humenné	59
Brezno	60
Sabinov	61
Stará Ľubovňa	62
Michalovce	63
Lučenec	64
Revúca	65
Košice - okolie	66
Levoča	67
Kysucké Nové Mesto	68
Šaľa	69
Stropkov	70
Spišská Nová Ves	71
Medzilaborce	72
Kežmarok	73
Veľký Krtíš	74
Gelnica	75
Rimavská Sobota	76
Trebišov	77
Rožňava	78
Sobrance	79

Výsledok postupu metódy poradí (december 1999) obsahuje predchádzajúca tabuľka. Čím je poradie nižšie, tým je lepšia situácia na trhu práce vzhľadom na sledované ukazovatele.

Metóda bodovacia

MESTO	PORADIE
Banská Bystrica	1
Trnava	2
Martin	3
Malacky	4
Piešťany	5
Košice IV	6
Bratislava I	7
Partizánske	8
Košice I	9
Bratislava III	10
Myjava	11
Bratislava II	12
Košice III	13
Považská Bystrica	14
Hlohovec	15
Michalovce	16
Púchov	17
Dolný Kubín	18
Nové Mesto nad Váhom	19
Senec	20
Bratislava IV	21
Košice II	22
Pezinok	23
Ilava	24
Bratislava V	25
Čadca	26
Poprad	27
Tvrdošín	28
Topoľčany	29
Žilina	30
Prievidza	31
Nitra	32
Trenčín	33
Detva	34
Liptovský Mikuláš	35
Vranov nad Topľou	36
Skalica	37
Bánovce nad Bebravou	38
Snina	39
Senica	40
Dunajská Streda	41
Kežmarok	42
Ružomberok	43
Turčianske Teplice	44
Galanta	45
Medzilaborce	46
Humenné	47
Spišská Nová Ves	48
Zlaté Moravce	49
Banská Štiavnica	50
Košice - okolie	51
Levice	52
Gelnica	53
Bytča	54
Levoča	55
Námestovo	56
Trebišov	57
Nové Zámky	58
Komárno	59
Prešov	60
Kysucké Nové Mesto	61
Rožňava	62
Stará Ľubovňa	63
Stropkov	64
Brezno	65
Svidník	66
Sabinov	67
Šaľa	68
Krupina	69
Sobrance	70
Žarnovica	71
Zvolen	72
Poltár	73
Žiar nad Hronom	74
Revúca	75
Bardejov	76
Lučenec	77
Rimavská Sobota	78
Veľký Krtíš	79

Najlepší je okres s poradím jedna. Najhorší okres má poradie 79.

Celkovo

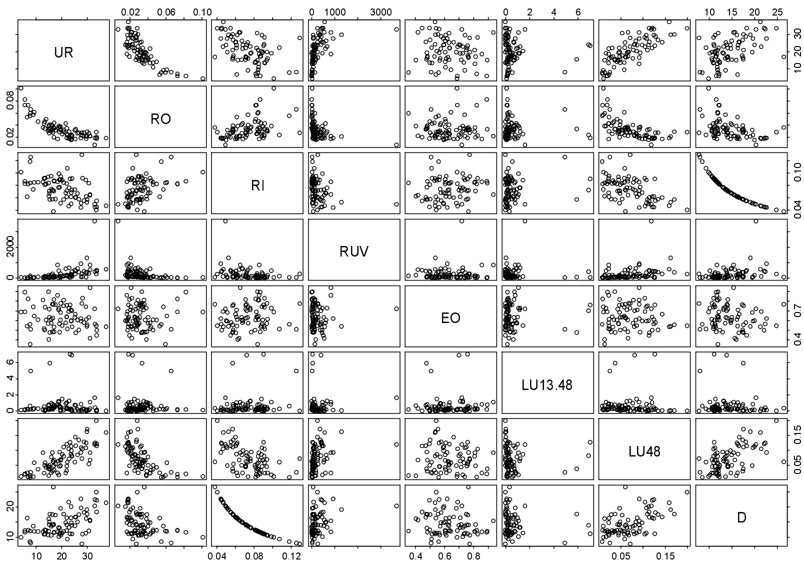
Pre každý okres (december 1999) sme dostali dve rôzne poradia, ako výsledok dvoch rôznych metód viackriteriálneho porovnávania. Postupy výpočtu poradí týmito metódami sú rozdielne, preto sa aj poradia navzájom líšia. Každá metóda má isté nedostatky, preto sa v praxi ako konečné poradie často počíta priemer z poradí, získaných všetkými metódami.

MESTO	PORADIE
Trnava	1
Banská Bystrica	2
Bratislava I	3
Bratislava II	4
Bratislava III	5
Piešťany	6
Malacky	7
Martin	8
Bratislava IV	9
Partizánske	10
Bratislava V	11
Hlohovec	12
Nové Mesto nad Váhom	13
Pezinok	14
Púchov	15
Myjava	16
Košice IV	17
Senec	18
Považská Bystrica	19
Liptovský Mikuláš	20
Trenčín	21
Ilava	22
Žilina	23
Dolný Kubín	24
Košice I	25
Prievidza	26
Poprad	27
Košice II	28
Topoľčany	29
Detva	30
Košice III	31
Tvrdošín	32
Čadca	33
Senica	34
Nitra	35
Galanta	36
Ružomberok	37
Michalovce	38
Skalica	39
Bánovce nad Bebravou	40
Vranov nad Topľou	41
Banská Štiavnica	42
Turčianske Teplice	43
Zvolen	44
Bytča	45
Dunajská Streda	46
Snina	47
Žiar nad Hronom	48
Humenné	49
Zlaté Moravce	50
Námestovo	51
Levice	52
Prešov	53
Svidník	54
Nové Zámky	55
Poltár	56
Komárno	57
Kežmarok	58
Košice - okolie	59
Medzilaborce	60
Spišská Nová Ves	61
Žarnovica	62
Levoča	63
Bardejov	64
Brezno	65
Krupina	66
Stará Ľubovňa	67
Gelnica	68
Sabinov	69
Kysucké Nové Mesto	70
Stropkov	71
Trebišov	72
Šaľa	73
Revúca	74
Rožňava	75
Lučenec	76
Sobrance	77
Veľký Krtíš	78
Rimavská Sobota	79

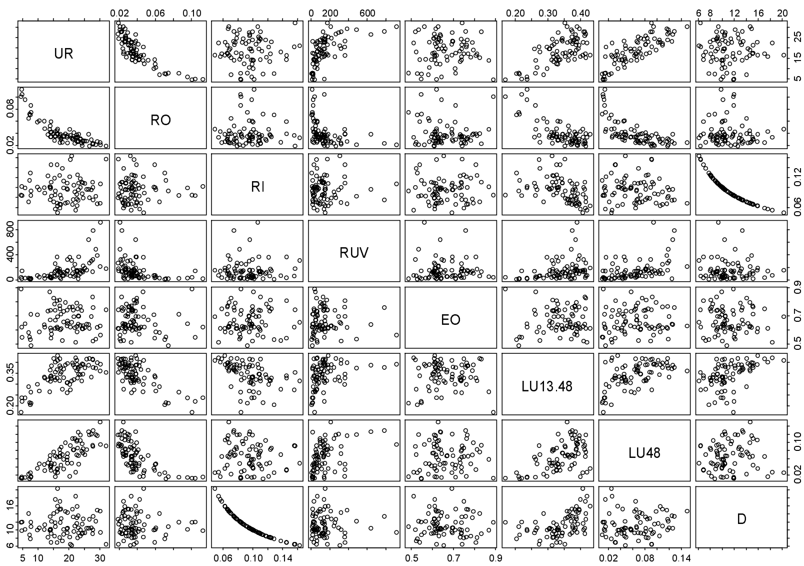
Grafické prostriedky priestorového porovnávania

Informácie z priestorového porovnania stavu nezamestnanosti v okresoch SR metódami viacrozmernej porovnávacej analýzy môžeme doplniť využitím grafických prostriedkov. Nasledujúci graf je vlastne „matica“ grafov všetkých dvojíc premenných na diagonále, pričom premenná na diagonále je závislou premennou vo všetkých grafoch v príslušnom stĺpci.

DECEMBER 1999:



DECEMBER 2000:



Z uvedených grafov môžeme o vzťahoch ukazovateľov stavu nezamestnanosti a trhu práce vysloviť niektoré zaujímavé závery. Z rozmiestnenia bodov je evidentná, skoro priama závislosť medzi UR a RO, medzi RI a D.a medzi RO a LU48.

DECEMBER 99

Faktorová analýza

- *Príprava a posúdenie vhodnosti dát*

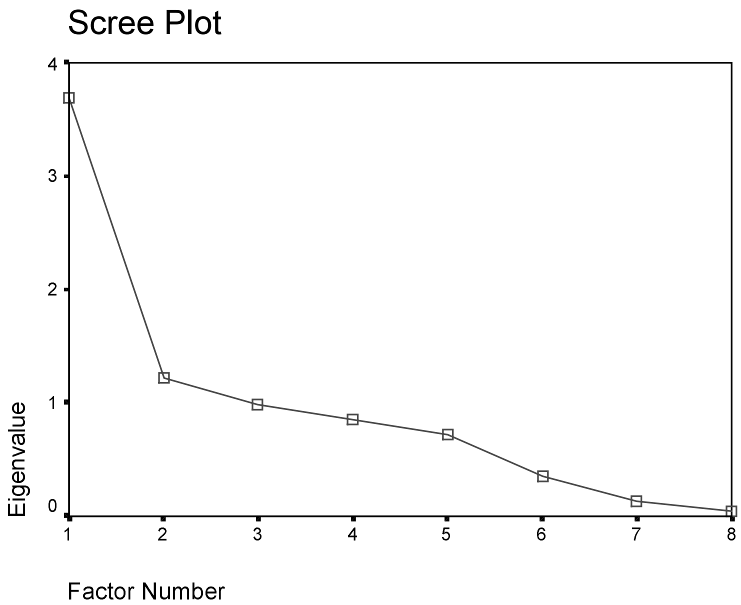
Základom pre zhodnotenie vhodnosti dát je analýza korelačnej matice. V prílohe (strana 50) som v korelačnej matici zvýraznil párové korelačné koeficienty korelácie, ktoré sú štatisticky významné a podľa absolútnej hodnoty indikujú významne silnú koreláciu. Korelačná matica naznačuje, že dáta sú vhodné pre použitie faktorovej analýzy.

Komplexne vhodnosť dát meria KMO index: Požaduje, sa aby index mal hodnoty vyššie ako 0,60. U našich dát je to číslo 0,781, takže dáta by sme mohli ohodnotiť ako náležite vhodné.

- *Odhad modelu faktorovej analýzy pred rotáciou*

Počet faktorov sme stanovili na základe testu, ktorý je dostupný pri MLE metóde odhadu faktorového modelu. Výsledok tohoto testu nám podporuje tri faktory. V scree plote je zlom pri dvojke, čo naznačuje, že jeden faktor by mohol stačiť. Lenže scree plot má až dva zlomy, čo nie je až taký častý jav. Z toho usudzujem, že by sme mali byť opatrnejší pri výbere počtu faktorov.

Keď použijeme pravidlo „eigenvalue-greater-than-one“[5] a scree plot, potom by sme mohli vybrať až tri faktory. Tretie vlastné číslo je menšie ako 1, ale jeho hodnota (0,979) nie je veľmi vzdialená od jednotky.



Výsledný nerotovaný model som odhadol metódou maximálnej vierohodnosti.

- *Rotácia*

V snahe získať ekonomicky vhodne interpretovateľnú faktorovú štruktúru som použil pravouhlú rotáciu QUARTIMAX. Tento typ rotácie zjednodušuje riadky – okresy, matice faktorovej štruktúry. Uvedená rotácia nám dávala hodnoty faktorových skóre, ktoré sú lepšie interpretovateľné.

Výsledky rotácie faktorov prehľadne zobrazuje nasledujúca tabuľka:

Premenná	Faktor 1	Faktor 2	Faktor 3	Komunalita	Unicita
UR	0.95	0.29	0.13	1.00	0.00
RO	-0.76	-0.23	-0.1	0.64	0.36
RI	-0.2	-0.89	-0.25	0.89	0.11
RUV	0.37	-0.01	0.93	1.00	0.00
EO	-0.07	-0.17	0.07	0.04	0.96
LU13.48	-0.02	-0.13	0.04	0.02	0.98
LU48	0.69	0.43	0.06	0.66	0.34
D	0.22	0.94	0.26	1.00	0.00
suma štvorcov fakt. Saturácií	2.19	2.04	1.03	5.26	2.74
Podiel na vysvetlenej variabilite (%)	**41.61**	**38.88**	**19.64**	**100**	
Podiel na celkovej variabilite (%)	**27.34**	**25.54**	**12.90**	**65.69**	**34.31**

Za významné som považoval faktorové saturácie, ktorých hodnota bola väčšia ako 0,62 (podľa tabuľky štatisticky významných faktorových saturácií).

Z tejto tabuľky možno získať nasledujúce zhrňujúce informácie o odhade rotovaného modelu faktorovej analýzy:

- celkovú variabilitu premennej po odhade modelu možno rozdeliť na časť vysvetlenú modelom – komunalitu a časť špecifickú – unicitu,
- existenciou troch skrytých faktorov možno u väčšiny premenných vysvetliť viac ako 60 % variability,
- výnimkou sú premenné EO a LU13-48,
- celkovo model vysvetľuje 65,69 % variability v dátach, čo možno pri analýze dát z trhu práce považovať za uspokojivé percento,
- z celkovej variability nemožno 34,31 % považovať za výsledok pôsobenia spoločných faktorov, je to teda variabilita vyvolaná pôsobením neidentifikovaných faktorov, alebo náhodných činiteľov,
- najviac k variabilite vysvetlenej modelom prispieva faktor 1, ktorého pôsobením možno vysvetliť 41,61 % spoločnej variability, čo je 27,34 % z celkovej variability.
- *Interpretácia výsledkov*

S prvým faktorom nám výrazne korelujú premenné UR, RO a LU48. Premenné UR a LU48 vystupujú s rovnakým znamienkom, avšak premenná RO koreluje s prvým faktorom so záporným znamienkom. Môžeme z toho usúdiť, že ten okres ktorý bude mať vysoké UR a LU48, bude mať nízke RO a naopak so zvyšujúcim RO sa bude znižovať UR a LU48. To je celkom logické nakoľko premenná RO vyjadruje podiel vyradených z evidencie k priemernému počtu nezamestnaných. Takže, ak sa zvýši hodnota RO, musí sa zákonite znížiť miera nezamestnanosti, nakoľko klesne počet nezamestnaných a klesne počet dlhodobo nezamestnaných (viac ako 48 mesiacov). Prvý faktor by mohol vyjadrovať vytváranie voľných pracovných miest prípadne vplyv podnikov a firiem na znižovanie nezamestnanosti poskytovaním práce. Najväčšie kladné hodnoty prvého faktora dosahujú okresy Rimavská Sobota, Vranov nad Topľou a Poltár. Naopak najväčšie záporné hodnoty dosahujú okresy Bratislava I, Bratislava IV a Bratislava III. Nulovú hodnotu má v prvom faktore okres Dunajská Streda. Z uvedeného by sa dalo usúdiť, že okresy, ktoré majú vysokú mieru nezamestnanosti (malé RO a veľké LU48), majú v prvom faktore vysoké kladné hodnoty, naopak okresy s malou mierou nezamestnanosti (veľké RO a malé LU48) majú v prvom faktore záporné hodnoty.

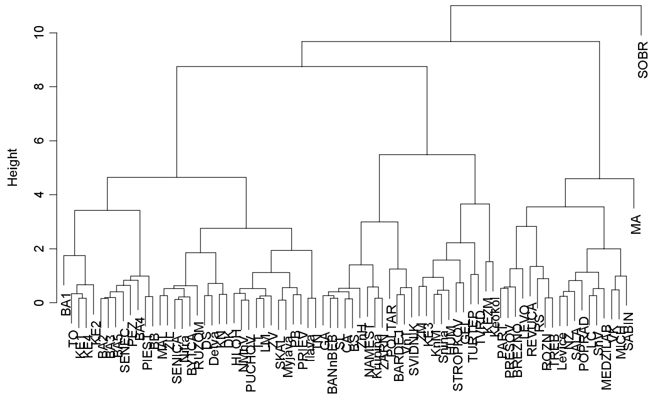
Druhý faktor by mohol prezentovať javy ktoré majú krátkodobú povahu, javy ktoré zapríčiňujú veľké, resp. malé množstvo novozaregistrovaných. Teda ak nejaký okres má RI väčšie ako druhý okres, znamená to že u prvého okresu prudšie vystúpil počet novozaregistrovaných ako u druhého. Druhý faktor je faktor, ktorý má dynamickú povahu a ktorý vyjadruje vplyv vecí a javov pôsobiacich v krátkej minulosti. Môže vyjadrovať sezónnosť, vplyvy úradov práce na zníženie nezamestanosti (verejnoprospešné práce) a pod. Je to súhra javov vplývajúci nielen, ale hlavne na posledný mesiac. Dalo sa očakávať že ukazovateľ D bude korelovať s druhým faktorom vzhľadom na jeho definíciu uvádzanú NÚP. Najväčšie kladné hodnoty prvého faktora nadobúdajú okresy Martin,Revúca a Prešov, najväčšie záporné hodnoty majú okresy Poltár, Ilava a Námestovo. Okresy s dlhou priemernou dobou nezamestnanosti majú vysoké kladné hodnoty v druhom faktore, naopak okresy s malou dobou nezamestnanosti nadobúdajú pri druhom faktore záporné hodnoty.

S tretím faktorom nám koreluje len premenná RUV, ktorá vyjadruje počet uchádzačov o zamestnanie pripadajúcich na jedno voľné pracovné miesto. Vysoký tretí faktor nám signalizuje nepriaznivú úroveň zamestnanosti v danom regióne, veľa nezamestnaných a málo pracovných miest. Najväčšie kladné hodnoty tretieho faktora dosahujú okresy Sobrance, Kežmarok a Košice okolie. V týchto okresoch je veľmi vysoký počet uchádzačov pripadajúcich na jedno voľné pracovné miesto.

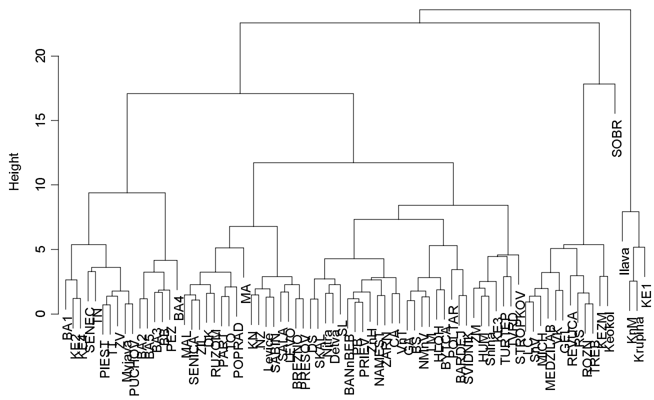
Zhluková analýza

Dendrogram sme spravili pomocou procedúry v programe S+ na **faktorových skóre**, ktoré sme dostali ako výstup z faktorovej analýzy. Pre porovnanie sme urobili dendrogram aj na pôvodných dátach. Z dendrogramu vidíme, v ktorých skupinách okresov je veľmi, v ktorých menej a v ktorých veľmi málo podobná situácia na trhu práce podľa zvolených ukazovateľov. Vyjadruje to úroveň, na ktorej sa okresy zhlukovali. Čím je úroveň zhlukovania vyššia, tým je menšia podobnosť okresov v danom zhluku. Na zhlukovanie sme použili Wardovu metódu zhlukovania. Na porovnanie jednotlivých metód som do prílohy zahrnul aj dendrogramy pre niektoré ďalšie metódy zhlukovania.

Dendrogram na faktorových skóre:



Dendrogram na pôvodných dátach:



V praxi sa veľmi neodporúča robiť zhlukovú analýzu na faktorových skóre, pretože môže dávať iné riešenie ako zhlukovanie na pôvodných dátach. Dendrogramy pre faktorové skóre a pre pôvodné dáta nie sú rovnaké, aj keď je tu badať určité podobnosti. Predovšetkým oba dendrogramy naznačujú, že okres Sobrance by mohol byť veľmi odlišný od ostatných okresov. Rozdielnosť medzi dendrogramami sa dá vysvetliť tým, že tri faktory, pomocou ktorých sme tvorili dendrogram, vysvetľujú len 65% celkovej variability pôvodných premenných. Počet zhlukov dostaneme tak, že posudzujeme vzdialenosť medzi zhlukmi v procese zhlukovania. Za významný považujeme okamih, keď zlúčime zhluky, ktorých vzdialenosť je výrazne väčšia ako vzdialenosť zhlukov ktoré sme zlúčili v predchádzajúcom kroku. Na základe tejto myšlienky, dendrogram pre pôvodné dáta naznačuje existenciu piatich zhlukov, kde jeden zhluk obsahuje len okres

Sobrance. Vytvorili sme 5 zhlukov a vypočítali sme priemerné hodnoty ukazovateľov a aj štandardné chyby priemerov:

```
***  Summary Statistics for data in:  dec99 ***

cluster.id1:1
           UR     RO     RI  RUV    EO LU13   LU48     D
Mean:    14.48 0.0428 0.0838  97 0.642 0.51 0.0460 12.21
SE Mean:  0.84 0.0029 0.0021  15 0.021 0.13 0.0043  0.29
------------------------------------------------------
cluster.id1:2
          UR    RO     RI  RUV    EO LU13   LU48    D
Mean:    25.9 0.0249 0.0537 257 0.588 0.53 0.1076 18.98
SE Mean:  1.3 0.0015 0.0015  38 0.022 0.25 0.0089  0.61
------------------------------------------------------
cluster.id1:3
          UR     RO     RI  RUV    EO LU13  LU48   D
Mean:    25.2 0.0255 0.0670 750 0.687 1.33 0.077 15.5
SE Mean:  1.3 0.0027 0.0041  88 0.044 0.71 0.014  1.1
------------------------------------------------------
cluster.id1:4
          UR     RO    RI   RUV    EO LU13  LU48  D
Mean:    26.6 0.0230 0.097 181 0.683 1.09 0.086 10.49
SE Mean:  1.2 0.0011 0.006  45 0.047 0.99 0.010  0.55
------------------------------------------------------
cluster.id1:5
         UR     RO    RI  RUV   EO LU13 LU48    D
Mean:     33 0.0068 0.049 3678 0.72  1.6 0.12   20
SE Mean: NA     NA    NA   NA   NA   NA   NA   NA
```

Príslušnosť jednotlivých okresov do zhlukov je v tabuľke v prílohe (strana 55). V prvom zhluku sú okresy, ktorých priemerné hodnoty ukazovateľov UR, LU48, RUV sú najmenšie v porovnaní s ostatnými zhlukmi. Druhý zhluk má priemernú hodnotu UR porovnateľnú so zhlukmi 3 a 4. Druhý zhluk však má najnižšiu priemernú hodnotu RI a najvyššiu hodnotu LU48 (ako zhluky 3 a 4). Zhluk 3 sa odlišuje od zhlukov 2 a 4 tým,. Že má výrazne vyššiu priemernú hodnotu RUV. Štvrtý zhluk má zase najvyššiu priemernú hodnotu RI, najmenšie RUV a D.V piatom zhluku je len jeden okres Sobrance, ktorý má veľmi vysoké UR a RUV.

DECEMBER 2000

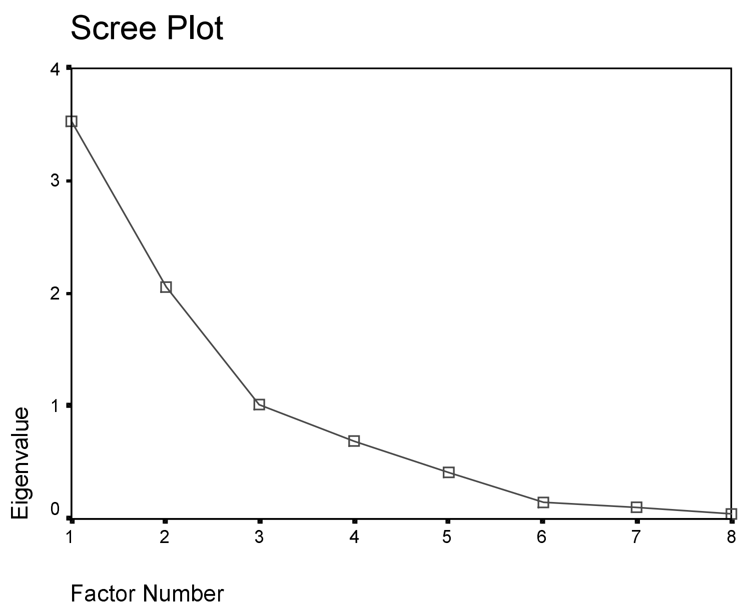
Faktorová analýza

- *Príprava a posúdenie vhodnosti dát*

Základom pre zhodnotenie vhodnosti dát je analýza korelačnej matice. V prílohe (strana 56) som v korelačnej matici zvýraznil párové korelačné koeficienty korelácie, ktoré sú štatisticky významné a podľa absolútnej hodnoty indikujú významne silnú koreláciu. Korelačná matica naznačuje, že dáta sú vhodné pre použite faktorovej analýzy.

Hodnota KMO indexu pre dáta z decembra roku 2000 je 0,696.

- *Odhad modelu faktorovej analýzy pred rotáciou*



Počet faktorov som stanovil na tri. Aj keď graf vlastných čísiel naznačuje, že by mali byť dva faktory, test pre počet faktorov (MLE) nám udáva, že treba použiť tri faktory. Výsledný nerotovaný model som odhadol metódou maximálnej vierohodnosti.

- *Rotácia*

Na analyzovanie som použil rotáciu PARSIMAX. Do analýz som nezahrnul okres Snina, nakoľko v decembri roku 2000 bolo 1553 uchádzačov o zamestnanie pripadajúcich na jedno voľné pracovné miesto, vďaka čomu bola negatívne ovplyvnená celková analýza (použili sme totiž metódu maximálnej vierohodnosti a tá je citlivá na tzv.

outliers). Tento negatívny trend mohol byť spôsobený hromadným prepúšťaným nejakého väčšieho podniku, čím sa výrazne zvýšil počet nezamestnaných prípadne klesol počet voľných miest. V nasledujúcej tabuľke sú prehľadne zhrnuté výsledky faktorovej analýzy aplikovanej na dáta za december roku 2000:

Premenná	Faktor 1	Faktor 2	Faktor 3	Komunalita	Unicita
UR	0.94	0.01	0.35	1.00	0.00
RO	-0.6	-0.09	-0.78	0.98	0.02
RI	-0.01	-0.95	0.09	0.91	0.09
RUV	0.62	-0.04	0	0.39	0.61
EO	-0.01	-0.03	0.12	0.02	0.98
LU13.48	0.53	0.5	0.47	0.75	0.25
LU48	0.76	0.08	0.28	0.66	0.34
D	0.02	0.99	-0.05	0.98	0.02
suma štvorcov fakt. Saturácií	2.49	2.15	1.06	5.69	2.31
podiel na vysvetlenej variabilite (%)	**43.74**	**37.81**	**18.56**	**100**	
podiel na celkovej variabilite (%)	**31.09**	**26.87**	**13.19**	**71.07**	**28.93**

Z tabuľky možno získať nasledujúce zhrňujúce informácie o odhade rotovaného modelu faktorovej analýzy:

- existenciou troch skrytých faktorov možno u väčšiny premenných vysvetliť viac ako 60 % variability,
- výnimkou sú premenné EO a RUV,
- celkovo model vysvetluje 71,07 % variability v dátach, čo možno pri analýze dát z trhu práce považovať za uspokojivé percento,
- z celkovej variability nemožno 28,93 % považovať za výsledok pôsobenia spoločných faktorov, je to teda variabilita vyvolaná pôsobením neidentifikovaných faktorov, alebo náhodných činiteľov,
- najviac k variabilite vysvetlenej modelom prispieva faktor 1, ktorého pôsobením možno vysvetliť 43,74 % spoločnej variability, čo je 31,09 % z celkovej variability.

Interpretácia výsledkov

Okresy, ktoré majú vysoké ukazovateľe UR, RUV a LU48, tie budú mať vysoké faktorové skóre u prvého faktora a naopak. Všetky tie tri premenné vstupujú s rovnakým znamienkom, z čoho sa dá usúdiť, že okres ktorý má vysokú UR, pravdepodobne bude mať vysoké RUV aj LU48. Miera nezamestnanosti, podiel uchádzačov pripadajúcich na jedno voľné pracovné miesto a podiel nezamestnaných dlhšie ako štyri roky k celkovému počtu dlhodobo nezamestaných sú javy spôsobené faktormi, ktoré sú dlhodobejšieho a vážneho charakteru. Vysoké skóre u prvého faktora vyjadruje vysoké problémy v danom okrese, majú málo voľných miest a veľa ľudí je dlhodobo nezamestnaných. Vysoké kladné hodnoty prvého faktora majú okresy Veľký Krtíš, Revúca a Spišská Nová Ves, naopak veľké záporné hodnoty nadobúdajú okresy Bratislava I, Ilava a Trenčín.

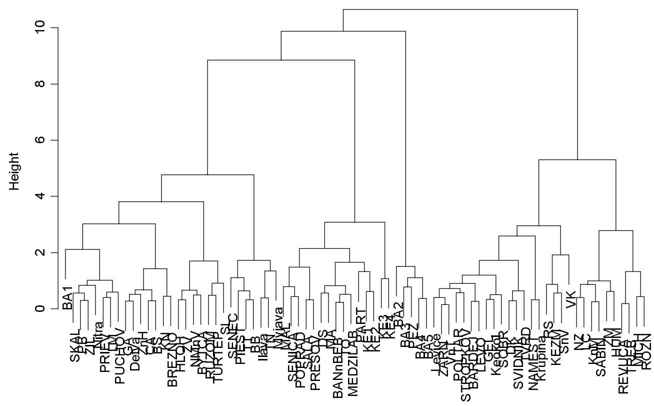
Druhý faktor by mohol prezentovať javy, ktoré majú krátkodobú povahu, javy ktoré zapríčiňujú veľké, resp. malé množstvo novozaregistrovaných. Teda ak nejaký okres má RI väčšie ako druhý okres, znamená to že u prvého okresu prudšie vystúpil počet novozaregistrovaných ako u druhého. Druhý faktor je faktor, ktorý má dynamickú povahu a ktorý vyjadruje vplyv vecí a javov pôsobiacich v krátkej minulosti. Môže vyjadrovať sezónnosť, vplyvy úradov práce na zníženie nezamestanosti (verejnoprospešné práce) a pod. Je to súhra javov vplývajúci nielen, ale hlavne na posledný mesiac. Dalo sa očakácať že ukazovateľ D bude korelovať s druhým faktorom vzhľadom na jeho definíciu uvádzanú NÚP. Najväčšie kladné hodnoty u daného faktora majú okresy Košice III, Košice II a Košice I. Najväčšie záporné hodnoty majú okresy Krupina, Veľký Krtíš a Námestovo.

Tretí faktor je faktor vyjadrujúci podiel vyradených z evidencie k priemernému počtu uchádzačov o zamestnanie. To, že RO nie je korelované s druhým faktorom, t.j. že nie je v skupine s RI, sa dá vysvetliť tým, že javy, ktoré stoja v pozadí toho, koľko ľudí dostane prácu, sú nezávislé od javov, ktoré stoja v pozadí toho, či niekto príde o prácu alebo nie. Najväčšie kladné hodnoty tretieho faktora majú okresy Bytča, Zvolen a Partizánske. Najväčšie záporné okresy Bratislava II, Bratislava V a Bratislava IV (tieto majú aj najväčšie hodnoty RO).

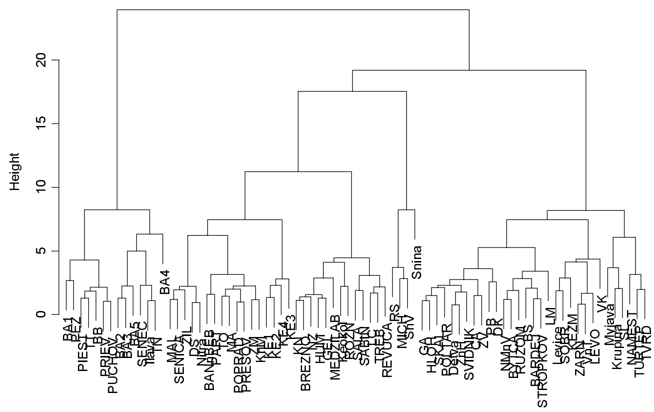
Zhluková analýza

Na zhlukovanie sme použili Wardovu metódu zhlukovania. Ako som už spomenul v mesiaci december 1999 v praxi sa veľmi neodporúča robiť zhlukovú analýzu na faktorových skóre, pretože môže dávať iné riešenie ako zhlukovanie na pôvodných dátach. Dendrogramy pre faktorové skóre a pre pôvodné dáta sú v konečnom dôsledku dosť podobné, nakoľko z obidvoch dendrogramov môžeme vidieť, že výsledne nám dávajú štyri zhluky. Počet zhlukov dostaneme tak, že posudzujeme vzdialenosť medzi zhlukmi v procese zhlukovania. Za významný považujeme okamih, keď zlúčime zhluky, ktorých vzdialenosť je výrazne väčšia ako vzdialenosť zhlukov, ktoré sme zlúčili v predchádzajúcom kroku.

Dendrogram na faktorových skóre:



Dendrogram na pôvodných dátach:



Počet zhlukov sme dostali tak, že sme posudzovali vzdialenosť medzi zhlukmi v procese zhlukovania a všímali sme si, kedy je táto vzdialenosť výrazne väčšia, ako vzdialenosť zhlukov, ktoré sme zlúčili v predchádzajúcom kroku. Vytvorili sme tak 4 zhluky a vypočítali sme priemerné hodnoty ukazovateľov a štandardné chyby priemerov:

```
***  Summary Statistics for data in:  dec99 ***

cluster.id:1
            UR     RO     RI RUV    EO  LU13   LU48     D
   Mean: 20.94 0.0318 0.0735 158 0.651 0.389 0.0807 13.98
SE Mean:  0.86 0.0016 0.0022  20 0.015 0.004 0.0059  0.43
------------------------------------------------------
cluster.id:2
           UR     RO    RI RUV   EO   LU13  LU48    D
   Mean: 26.1 0.0298 0.099 973 0.65 0.3825 0.100 10.4
SE Mean:  2.2 0.0046 0.009 201 0.04 0.0032 0.018  1.1
------------------------------------------------------
cluster.id:3
           UR     RO     RI  RUV    EO  LU13   LU48     D
   Mean: 8.29 0.0750 0.0958 38.2 0.623 0.256 0.0243 10.64
SE Mean: 0.81 0.0058 0.0038  9.4 0.027 0.013 0.0036  0.41
------------------------------------------------------
cluster.id:4
            UR     RO     RI RUV    EO   LU13   LU48    D
   Mean: 20.12 0.0340 0.1153 145 0.701 0.3284 0.0643 8.91
SE Mean:  0.84 0.0014 0.0037  18 0.015 0.0062 0.0047 0.26
```

Príslušnosť jednotlivých okresov do zhlukov je v tabuľke v prílohe (strana 62). V treťom zhluku sú okresy, ktorých priemerné hodnoty ukazovateľov UR, LU13.48, LU48 a RUV sú najmenšie v porovnaní s ostatnými zhlukmi. Priemerná hodnota RO je v tomto zhluku najvyššia. Patria sem okresy, ktoré majú situáciu na trhu práce najlepšiu. Prvý a štvrtý zhluk má priemernú hodnotu UR porovnateľnú, ale vyššiu ako v tretí zhluk. Prvý zhluk má však v porovnaní so štvrtým výrazne menšiu hodnotu RI a vyššiu hodnotu LU13.48, LU48 a D. V druhom zhluku sú okresy s najvyššou hodnotou UR, RI, RUV, LU48 a s najmenšou priemernou hodnotou ukazovateľa RO. Zjavne sem patria okresy s najväčšími problémami, čo sa týka krátkodobej i dlhodobej povahy.

Záver

Popísali sme tri štatistické metódy: metóda viackriteriálneho porovnávania, zhluková a faktorová metóda. Aplikovali sme ich následne s cieľom porovnať okresy na Slovensku z pohľadu nezamestnanosti. Výhodou týchto metód je že pri porovnávaní berú do úvahy viacero ukazovateľov súčasne a teda môžu byť objektívnejšie. Z uvedených analýz nám vyplýva, že v decembri roku 1999 boli na tom okresy Rimavská Sobota, Sobrance a Revúca čo sa týka miery nezamestnanosti a rôznych ďaľších kritérií horšie v porovnaní s okresmi iných krajov. V decembri roku 2000 to boli okresy Veľký Krtíš, Revúca a časti Košíc. Jednotlivým konkrétnym prípadom sa venujem v interpretácii výsledkov.

Metódy viackriteriálneho porovnávania (ktoré sme urobili pre rok 1999) nám dávajú podobné výsledky, ako metódy faktorovej analýzy.

Uvedené metódy (faktorová a zhluková analýza) sa môžu používať nielen na analýzy nezamestnanosti, ale aj v úplne odlišných oblastiach ako napríklad v psychológii a finančnej analýze, kde sa môžeme stretnúť s veľkým množstvom ukazovateľov súčasne.

Povaha faktorovej analýzy je skôr heuristická ako overovacia. Úspešné použitie faktorovej analýzy vyžaduje znalosti v aplikačnej oblasti, rešpektovanie predpokladov metódy a určité praktické skúsenosti a praktickým uplatnením.

Uplatnenie metód zhlukovej analýzy vedie k priaznivým výsledkom predovšetkým tam, kde sa študovaný objekt reálne rozpadá do tried, tj. objekty majú tendenciu zoskupovať sa do prirodzených zhlukov.

Použitá literatúra

1. **Analýzy trhu práce, Projekt Phare No. T – 9108 – 41/06/28 WP 5**
2. **SPSS for Windows, Professional Statistics, release 6.0, Marija J.,SPSS Inc., 1993**
3. **Multivariate Data Analysis, Hair,J.F, Anderson R.E., Tatham, R.L., Black, W.C, fourth edition, prentice hall, englewood cliffs, 1995**
4. **Vícerozmerné štatistické metódy s aplikacemi, Hebák,P., Hustopecký,J., SNTL/ALFA, Praha 1987**
5. **Applied multivariate techniques, Sharma,S., J. Wiley & Sons Inc., New York 1996**
6. **http://www.nup.sk**

Príloha

Na nasledujúcich stranách prezentujem výsledky programu S+ pre faktorovú a zhlukovú analýzu za obidva analyzované mesiace. Pre porovnanie uvediem nielen použité metódy, ale výsledky ďaľších metód. Nebudem uvádzať všetky ktoré som spúšťal. Uvedené metódy dávali rovnaké, ale neinterpretovateľné výsledky.

DECEMBER 1999

Korelačná matica

```
***  Correlation for data in:  DECEMBER99U ***

                 UR          RO          RI         RUV          EO
     UR  1.00000000 -0.80208568 -0.4816438  0.47117659 -0.10847194
     RO -0.80208568  1.00000000  0.4176872 -0.37391692  0.01412235
     RI -0.48164381  0.41768721  1.0000000 -0.29438383  0.14487951
    RUV  0.47117659 -0.37391692 -0.2943838  1.00000000  0.03869559
     EO -0.10847194  0.01412235  0.1448795  0.03869559  1.00000000
LU13.48 -0.04738072 -0.04322577  0.1216383  0.03643553  0.02197557
   LU48  0.78417507 -0.62386865 -0.5351577  0.30966495 -0.09085211
      D  0.51437951 -0.40778508 -0.9482340  0.30986457 -0.16054800
            LU13.48        LU48          D
     UR -0.04738072  0.78417507  0.5143795
     RO -0.04322577 -0.62386865 -0.4077851
     RI  0.12163827 -0.53515768 -0.9482340
    RUV  0.03643553  0.30966495  0.3098646
     EO  0.02197557 -0.09085211 -0.1605480
LU13.48  1.00000000 -0.03613794 -0.1108384
   LU48 -0.03613794  1.00000000  0.5688330
      D -0.11083844  0.56883297  1.0000000
```

Najprv nasleduje výstup z faktorovej analýzy za december roku 1999 pred rotáciou.

```
      *** Factor Analysis ***
Sums of squares of loadings:
 Factor1   Factor2   Factor3
 3.51071 0.9513263 0.7976492

The number of variables is 8 and the number of observations is 79

Test of the hypothesis that 3 factors are sufficient
versus the alternative that more are required:
The chi square statistic is 6.04 on 7 degrees of freedom.
The p-value is 0.535

Component names:

 "loadings" "uniquenesses" "correlation" "criteria" "factors" "dof" "method"

 "center" "scale" "n.obs" "scores" "call"

Call:
factanal(x =  ~ UR + RO + RI + RUV + EO + LU13.48 + LU48 + D, factors = 3,
       method = "mle", data = DECEMBER99U, scores = T, type = "regression",
       rotation = "none", na.action = na.omit, control = list(iter.max = 20,
       unique.tol = 0.0001))

Importance of factors:
                 Factor1   Factor2    Factor3
   SS loadings 3.5107102 0.9513263 0.79764918
Proportion Var 0.4388388 0.1189158 0.09970615
Cumulative Var 0.4388388 0.5577546 0.65746072
```

```
The degrees of freedom for the model is 7.

Uniquenesses:
     UR        RO        RI           RUV       EO   LU13.48      LU48      D
 4e-010 0.3566109 0.1007888 0.00003581632 0.960506 0.9818249 0.3405488 4e-010

Loadings:
        Factor1 Factor2 Factor3
     UR  0.870   0.493
     RO -0.695  -0.400
     RI -0.822   0.473
    RUV  0.449   0.164   0.878
     EO -0.155           0.113
LU13.48
   LU48  0.777   0.219
      D  0.870  -0.493
```

Naledujúci výstup je faktorová analýza decembra roku 1999 pri rotácii QUARTIMAX:

```
      *** Factor Analysis ***
Sums of squares of loadings:
  Factor1  Factor2  Factor3
 2.181864 2.048983 1.028839
The number of variables is 8 and the number of observations is 79

Test of the hypothesis that 3 factors are sufficient
versus the alternative that more are required:
The chi square statistic is 6.04 on 7 degrees of freedom.
The p-value is 0.535

Component names:

 "loadings" "uniquenesses" "correlation" "criteria" "factors" "dof" "method"

 "center" "scale" "n.obs" "scores" "call"

Call:
factanal(x =  ~ UR + RO + RI + RUV + EO + LU13.48 + LU48 + D, factors = 3,
       method = "mle", data = DECEMBER99U, scores = T, type = "regression",
       rotation = "quartimax", na.action = na.omit, control = list(iter.max =
20, unique.tol = 0.0001))

Importance of factors:
                Factor1   Factor2   Factor3
   SS loadings 2.181864 2.0489828 1.0288392
Proportion Var 0.272733 0.2561229 0.1286049
Cumulative Var 0.272733 0.5288558 0.6574607

The degrees of freedom for the model is 7.

Uniquenesses:
     UR        RO        RI           RUV       EO   LU13.48      LU48      D
 4e-010 0.3566109 0.1007888 0.00003581632 0.960506 0.9818249 0.3405488 4e-010

Loadings:
        Factor1 Factor2 Factor3
     UR  0.948   0.291   0.132
     RO -0.762  -0.229  -0.101
     RI -0.200  -0.893  -0.248
    RUV  0.371           0.929
     EO         -0.173
LU13.48         -0.126
   LU48  0.687   0.428
      D  0.218   0.941   0.258
```

TABUĽKA FAKTOROVÝCH SKÓRE

	Faktor 1	Faktor 2	Faktor 3
Bratislava I	-2.18	0.9	0.26
Bratislava II	-1.88	-0.39	0.13
Bratislava III	-1.93	-0.24	0.16
Bratislava IV	-1.95	-0.85	0.2
Bratislava V	-1.75	-0.24	0.12
Malacky	-0.77	0.37	-0.21
Pezinok	-1.57	0.12	0.04
Senec	-1.62	-0.41	0.35
Dunajská Streda	0	-0.31	-0.36
Galanta	0.29	-0.94	-0.63
Hlohovec	-0.54	-0.47	-0.35
Piešťany	-1.14	-0.46	-0.01
Senica	-0.47	0.06	-0.31
Skalica	-0.18	-0.64	-0.39
Trnava	-0.71	-0.38	-0.22
Bánovce nad Bebravou	0.15	-0.78	-0.43
Ilava	-1.22	-1.45	-0.1
Myjava	-0.17	-0.65	-0.49
Nové Mesto nad Váhom	-0.66	-0.73	-0.23
Partizánske	-0.52	1.56	-0.34
Považská Bystrica	-0.29	-0.97	-0.09
Prievidza	-0.45	-0.84	-0.32
Púchov	-0.7	-0.71	-0.31
Trenčín	-1.3	-1.31	-0.04
Komárno	0.52	0.07	-0.48
Levice	0.55	0.52	-0.24
Nitra	-0.53	-0.05	-0.29
Nové Zámky	0.34	0.46	-0.16
Šaľa	0.1	0.55	-0.31
Topoľčany	-0.63	0.93	-0.15
Zlaté Moravce	0.45	-0.67	0.75
Bytča	-0.44	0.03	-0.12
Čadca	0.48	-0.65	-0.39
Dolný Kubín	0.31	-0.1	-0.32
Kysucké Nové Mesto	0.59	-0.33	0.08
Liptovský Mikuláš	-0.84	-0.48	-0.24
Martin	-1.58	3.53	0.19
Námestovo	0.78	-1.39	-0.68
Ružomberok	-0.53	-0.09	0.22
Turčianske Teplice	-0.11	-1.11	1.37

	Faktor 1	Faktor 2	Faktor 3
Tvrdošín	0.17	-1.32	0.39
Žilina	-0.7	0.11	-0.19
Banská Bystrica	-1.36	-0.4	-0.01
Banská Štiavnica	0.33	-0.67	-0.41
Brezno	0.03	1.53	-0.18
Detva	-0.06	-0.19	-0.59
Krupina	0.8	-0.86	-0.87
Lučenec	0.68	0.63	-0.01
Poltár	1.75	-2.02	-0.73
Revúca	1.04	2.65	-0.41
Rimavská Sobota	1.87	1.41	-0.03
Veľký Krtíš	1.62	0.66	-0.09
Zvolen	-0.81	-0.39	-0.01
Žarnovica	0.91	-0.94	-0.77
Žiar nad Hronom	0.19	-0.91	-0.1
Bardejov	1.65	-1.11	-0.5
Humenné	0.79	-0.2	0.52
Kežmarok	0.86	0.46	2.08
Levoča	0.64	1.42	-0.69
Medzilaborce	0.58	0.35	0.22
Poprad	-0.19	0.94	-0.27
Prešov	-0.25	2.02	-0.29
Sabinov	1.27	0.41	-0.98
Snina	0.76	-0.48	0.3
Stará Ľubovňa	0.3	-0.77	-0.53
Stropkov	1.45	-0.31	0.76
Svidník	1.52	-1.02	-1.03
Vranov nad Topľou	1.77	-0.9	-0.67
Gelnica	1.47	-0.05	0.05
Košice I	-0.97	1.07	-0.11
Košice II	-0.92	0.57	0.16
Košice III	0.13	-0.16	0.72
Košice IV	-0.82	1.02	-0.16
Košice – okolie	0.24	1.73	1.48
Michalovce	1.15	0.62	-0.22
Rožňava	1.16	1.88	-0.11
Sobrance	0.86	-0.73	7.65
Spišská Nová Ves	0.91	0.75	0.02
Trebišov	1.28	1.76	-0.05

KORELAČNÁ MATICA

	UR	RO	RI	RUV	EO	LU13.48	LU48	D
UR	1	-0.8	-0.48	0.47	-0.11	-0.05	0.78	0.51
RO	-0.8	1	0.42	-0.37	0.01	-0.04	-0.62	-0.41
RI	-0.48	0.42	1	-0.29	0.14	0.12	-0.54	-0.95
RUV	0.47	-0.37	-0.29	1	0.04	0.04	0.31	0.31
EO	-0.11	0.01	0.14	0.04	1	0.02	-0.09	-0.16
LU13.48	-0.05	-0.04	0.12	0.04	0.02	1	-0.04	-0.11
LU48	0.78	-0.62	-0.54	0.31	-0.09	-0.04	1	0.57
D	0.51	-0.41	-0.95	0.31	-0.16	-0.11	0.57	1

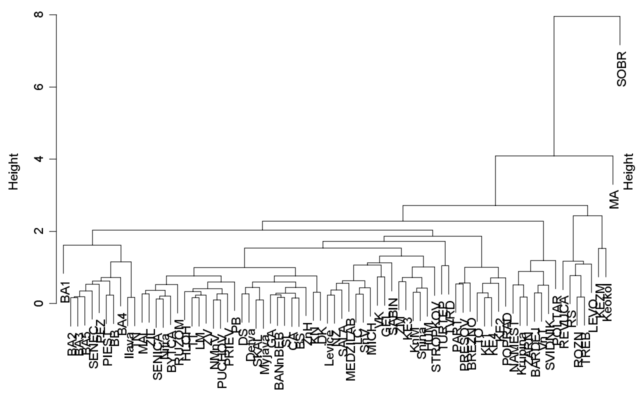
Ďalej nasleduje pre porovnanie ortogonálna rotácia VARIMAX, a kosouhlá rotácia COVARIMIN:

```
       *** Factor Analysis ***
Sums of squares of loadings:
  Factor1  Factor2  Factor3
 2.245223 1.929184 1.085279

The number of variables is 8 and the number of observations is 79

Test of the hypothesis that 3 factors are sufficient
versus the alternative that more are required:
The chi square statistic is 6.04 on 7 degrees of freedom.
The p-value is 0.535

Component names:

 "loadings" "uniquenesses" "correlation" "criteria" "factors" "dof" "method"

 "center" "scale" "n.obs" "scores" "call"

Call:
factanal(x =  ~ UR + RO + RI + RUV + EO + LU13.48 + LU48 + D, factors = 3,
       method = "mle", data = DECEMBER99U, scores = T, type = "regression",
       rotation = "varimax", na.action = na.omit, control = list(iter.max =
       20, unique.tol = 0.0001))

Importance of factors:
                 Factor1   Factor2   Factor3
   SS loadings 2.2452228 1.9291841 1.0852788
Proportion Var 0.2806528 0.2411480 0.1356599
Cumulative Var 0.2806528 0.5218009 0.6574607

The degrees of freedom for the model is 7.

Uniquenesses:
     UR        RO        RI            RUV       EO   LU13.48      LU48      D
 4e-010 0.3566109 0.1007888 0.00003581632 0.960506 0.9818249 0.3405488 4e-010

Loadings:
        Factor1 Factor2 Factor3
     UR  0.956   0.230   0.184
     RO -0.768  -0.180  -0.142
     RI -0.246  -0.886  -0.232
    RUV  0.311           0.950
     EO         -0.165
LU13.48         -0.123
   LU48  0.710   0.382
      D  0.267   0.933   0.241
       *** Factor Analysis ***
Sums of squares of loadings:
 Factor1  Factor2   Factor3
 2.08828 1.875503 0.9783502

The number of variables is 8 and the number of observations is 79

Test of the hypothesis that 3 factors are sufficient
versus the alternative that more are required:
The chi square statistic is 6.04 on 7 degrees of freedom.
The p-value is 0.535

Component names:
```

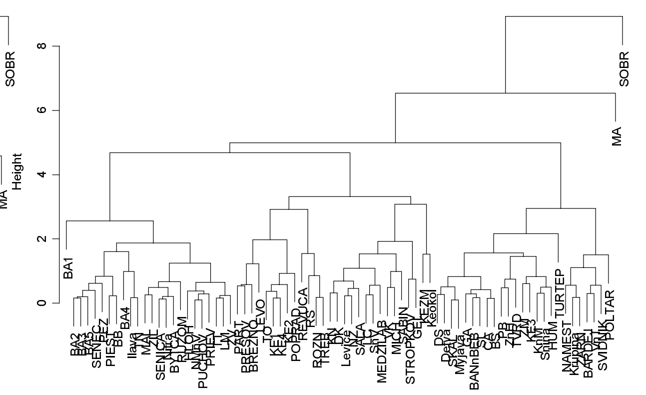
```
 "loadings" "uniquenesses" "correlation" "criteria" "factors" "dof" "method"

 "center" "scale" "n.obs" "scores" "call"

Call:
factanal(x =  ~ UR + RO + RI + RUV + EO + LU13.48 + LU48 + D, factors = 3,
       method = "mle", data = DECEMBER99U, scores = T, type = "regression",
       rotation = "covarimin", na.action = na.omit, control = list(iter.max
        = 20, unique.tol = 0.0001))

Importance of factors:
                   Factor1   Factor2   Factor3
     SS loadings 2.088280 1.8755025 0.9783502
  Variable Index 0.261035 0.2344378 0.1222938
Cumulative Index 0.261035 0.4954728 0.6177666

The degrees of freedom for the model is 7.

Uniquenesses:
     UR        RO        RI           RUV       EO   LU13.48      LU48      D
 4e-010 0.3566109 0.1007888 0.00003581632 0.960506 0.9818249 0.3405488 4e-010

Loadings:
        Factor1 Factor2 Factor3
     UR  0.972
     RO -0.783
     RI         -0.919  -0.182
    RUV  0.212   0.104   0.941
     EO         -0.138
LU13.48         -0.113
   LU48  0.686   0.234
      D          0.965   0.188

Component/Factor Correlations:
        Factor1 Factor2 Factor3
Factor1  1.000  -0.408   0.145
Factor2 -0.408   1.000   0.091
Factor3  0.145   0.091   1.000
```

Teraz nasledujú výsledky zhlukovania – dendrogramy pre rôzne metódy zhlukovania v uvedenom poradí: AVERAGE, COMPLETE, SINGLE, WEIGHTED, pri použití euklidovskej vzdialenosti.

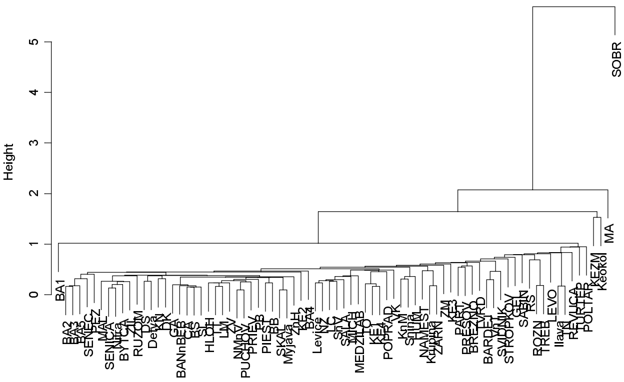
AVERAGE



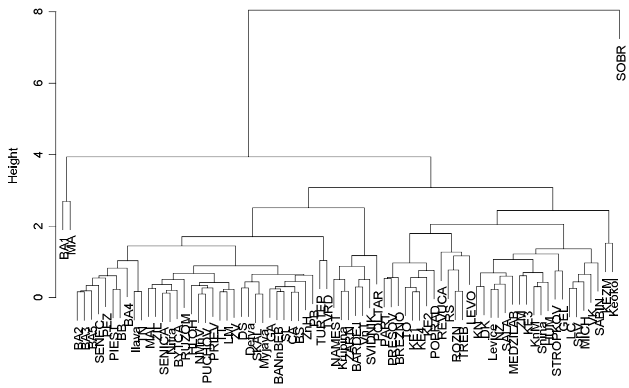
COMPLETE



SINGLE



WEIGHTED



Tabuľka príslušnosti do zhlukov:

Okres	Cluster
Bratislava I	1
Bratislava II	1
Bratislava III	1
Bratislava IV	1
Bratislava V	1
Malacky	1
Pezinok	1
Senec	1
Dunajská Streda	1
Galanta	1
Hlohovec	1
Pieštany	1
Senica	1
Skalica	1
Trnava	1
Bánovce n/Bebravou	1
Ilava	1
Myjava	1
Nové Mesto n/Váhom	1
Partizánske	2
Považská Bystrica	1
Prievidza	1
Púchov	1
Trencín	1
Komárno	1
Levice	2
Nitra	1
Nové Zámky	2
Šala	2
Topolcany	2
Zlaté Moravce	3
Bytca	1
Cadca	1
Dolný Kubín	1
Kysucké Nové Mesto	3
Liptovský Mikuláš	1
Martin	2
Námestovo	4
Ružomberok	1
Turcianske Teplice	3
Tvrdošín	1
Žilina	1
Banská Bystrica	1
Banská Štiavnica	1
Brezno	2
Detva	1
Krupina	4
Lucenec	2
Poltár	4
Revúca	2
Rimavská Sobota	2
Velký Krtíš	2
Zvolen	1
Žarnovica	4
Žiar nad Hronom	1
Bardejov	4
Humenné	3
Kežmarok	3
Levoca	2
Medzilaborce	2
Poprad	2
Prešov	2
Sabinov	2
Snina	3
Stará Lubovna	1
Stropkov	3
Svidník	4
Vranov nad Toplou	4
Gelnica	2
Košice I	2
Košice II	1
Košice III	3
Košice IV	2
Košice - okolie	3
Michalovce	2
Rožnava	2
Sobrance	5
Spišská Nová Ves	2
Trebišov	2

DECEMBER 2000

Korelačná matica:

```
***  Correlation for data in:  DECEMBER00 ***

                 UR          RO          RI          RUV          EO
     UR  1.000000000 -0.83782314  0.01801398  0.58329286  0.03624820
     RO -0.837823137  1.00000000  0.01844927 -0.37424450 -0.08721573
     RI  0.018013976  0.01844927  1.00000000  0.02144912  0.02880846
    RUV  0.583292855 -0.37424450  0.02144912  1.00000000 -0.01535480
     EO  0.036248199 -0.08721573  0.02880846 -0.01535480  1.00000000
LU13.48  0.668374146 -0.73076838 -0.42923109  0.32309061  0.03329292
   LU48  0.806458268 -0.68217930 -0.10850435  0.43147508  0.06784023
      D  0.009380693 -0.06174806 -0.94690317 -0.02616004 -0.03181031
           LU13.48        LU48           D
     UR  0.66837415  0.80645827  0.009380693
     RO -0.73076838 -0.68217930 -0.061748064
     RI -0.42923109 -0.10850435 -0.946903166
    RUV  0.32309061  0.43147508 -0.026160037
     EO  0.03329292  0.06784023 -0.031810313
LU13.48  1.00000000  0.54349843  0.482646349
   LU48  0.54349843  1.00000000  0.071060841
      D  0.48264635  0.07106084  1.000000000
```

Najprv nasleduje výstup z faktorovej analýzy za december roku 2000 pred rotáciou:

```
       *** Factor Analysis ***
Sums of squares of loadings:
  Factor1  Factor2   Factor3
 3.141004 2.139263 0.4046237
The number of variables is 8 and the number of observations is 78
Test of the hypothesis that 3 factors are sufficient
versus the alternative that more are required:
The chi square statistic is 10.42 on 7 degrees of freedom.
The p-value is 0.166
Component names:

 "loadings" "uniquenesses" "correlation" "criteria" "factors" "dof" "method"

 "center" "scale" "n.obs" "scores" "call"
Call:
factanal(x =  ~ UR + RO + RI + RUV + EO + LU13.48 + LU48 + D, factors = 3,
       method = "mle", data = DECEMBER00U, scores = T, type =
       "regression", rotation = "none", na.action = na.omit, control = list(
       iter.max = 20, unique.tol = 0.0001))
Importance of factors:
                 Factor1   Factor2    Factor3
   SS loadings 3.1410035 2.1392630 0.40462374
Proportion Var 0.3926254 0.2674079 0.05057797
Cumulative Var 0.3926254 0.6600333 0.71061129

The degrees of freedom for the model is 7.
Uniquenesses:
     UR         RO         RI       RUV        EO   LU13.48      LU48
 4e-010 0.02412681 0.08868414 0.6120188 0.9851462 0.2467416 0.3445094
         D
 0.01390979
Loadings:
        Factor1 Factor2 Factor3
     UR  1.000
     RO -0.838           0.519
     RI         -0.953
    RUV  0.583           0.216
```

```
     EO                 -0.113
LU13.48  0.668   0.484  -0.269
   LU48  0.806
      D          0.993
```

Nasledujúci výstup je faktorová analýza decembra roku 2000 pri rotácii PARSIMAX.

```
       *** Factor Analysis ***
Sums of squares of loadings:
  Factor1  Factor2  Factor3
 2.484011 2.150241 1.050638

The number of variables is 8 and the number of observations is 78

Test of the hypothesis that 3 factors are sufficient
versus the alternative that more are required:
The chi square statistic is 10.42 on 7 degrees of freedom.
The p-value is 0.166

Component names:

 "loadings" "uniquenesses" "correlation" "criteria" "factors" "dof" "method"

 "center" "scale" "n.obs" "scores" "call"
Call:
factanal(x =  ~ UR + RO + RI + RUV + EO + LU13.48 + LU48 + D, factors = 3,
       method = "mle", data = DECEMBER00U, scores = T, type =
       "regression", rotation = "parsimax", na.action = na.omit, control =
       list(iter.max = 20, unique.tol = 0.0001))
Importance of factors:
                 Factor1   Factor2   Factor3
   SS loadings 2.4840113 2.1502413 1.0506377
Proportion Var 0.3105014 0.2687802 0.1313297
Cumulative Var 0.3105014 0.5792816 0.7106113
The degrees of freedom for the model is 7.


Uniquenesses:
     UR         RO         RI       RUV        EO   LU13.48      LU48
 4e-010 0.02412681 0.08868414 0.6120188 0.9851462 0.2467416 0.3445094
         D
 0.01390979
Loadings:
        Factor1 Factor2 Factor3
     UR  0.936           0.351
     RO -0.603          -0.778
     RI         -0.950
    RUV  0.622
     EO                  0.119
LU13.48  0.534   0.498   0.468
   LU48  0.756           0.280
      D          0.992
```

TABUĽKA FAKTOROVÝCH SKÓRE

	Faktor 1	Faktor 2	Faktor 3
Bratislava I	-1.74	0.97	-0.39
Bratislava II	-0.85	-0.65	-3.68
Bratislava III	-1.11	-0.48	-2.53
Bratislava IV	-1.18	0.07	-2.9
Bratislava V	-1.01	0.1	-3.21
Malacky	-0.64	0.73	-0.31
Pezinok	-0.97	0.15	-2.11
Senec	-1.29	-0.19	-1.22
Dunajská Streda	-0.45	0.62	0.72
Galanta	-0.26	-0.33	0.59
Hlohovec	-1.2	-0.32	1.26
Piešťany	-1.38	-0.39	-0.08
Senica	-0.36	1.15	-0.1
Skalica	-0.83	-0.05	0.94
Trnava	-0.86	-0.37	-0.51
Bánovce nad Bebravou	-0.52	1.38	0.86
Ilava	-1.53	-1.24	-0.66
Myjava	-0.49	-1.57	-0.55
Nové Mesto nad Váhom	-1.3	-0.98	1.2
Partizánske	-0.73	2.02	1.26
Považská Bystrica	-0.78	0.26	0.76
Prievidza	-0.95	-0.28	0.6
Púchov	-1.17	0.09	0.38
Trenčín	-1.52	-0.87	-0.79
Komárno	0.62	0.14	0.8
Levice	0.81	-0.39	0.06
Nitra	-0.32	0.25	0.28
Nové Zámky	0.92	0.52	0.11
Šaľa	0.54	1.54	0.28
Topoľčany	-0.49	1.33	0.63
Zlaté Moravce	0.8	0.58	-0.45
Bytča	-1.13	-0.85	1.65
Čadca	-0.25	-0.75	0.57
Dolný Kubín	0.92	-0.73	-0.47
Kysucké Nové Mesto	0.33	0.71	0.3
Liptovský Mikuláš	-0.85	-0.42	0.23
Martin	-0.78	1.04	0.56
Námestovo	0.17	-1.86	0.58
Ružomberok	-0.52	-1.05	0.73
Turčianske Teplice	-0.64	-1.52	0.72

	Faktor 1	Faktor 2	Faktor 3
Tvrdošín	0.54	-1.1	-0.68
Žilina	-0.88	0.39	0.51
Banská Bystrica	-1.14	-0.25	-0.44
Banská Štiavnica	-0.01	-0.69	0.42
Brezno	0.21	-0.01	0.89
Detva	-0.03	-0.31	0.44
Krupina	0.21	-1.77	0.27
Lučenec	1.12	0.5	0.46
Poltár	0.57	-0.2	0.43
Revúca	2.12	1.06	-0.73
Rimavská Sobota	1.94	-0.68	-0.19
Veľký Krtíš	2.18	-1.84	-0.02
Zvolen	-1.17	-0.64	1.26
Žarnovica	0.87	-0.71	0.21
Žiar nad Hronom	0.1	-0.26	0.15
Bardejov	0.67	-1.15	0.05
Humenné	0.65	0.16	-0.02
Kežmarok	1.58	-0.3	-0.48
Levoča	0.76	-0.94	0.19
Medzilaborce	-0.22	1.11	0.99
Poprad	0.07	1.15	-0.08
Prešov	0.31	1.33	0.24
Sabinov	0.57	0.79	0.37
Stará Ľubovňa	0.68	-0.74	0.57
Stropkov	0.66	-0.6	-0.45
Svidník	1.01	-0.67	0.16
Vranov nad Topľou	1.12	-0.26	0.09
Gelnica	-0.51	1.62	0.34
Košice I	-0.48	2.11	-0.03
Košice II	0.67	2.3	-0.3
Košice III	-0.16	2.91	-0.6
Košice IV	1.41	-0.4	0.17
Košice - okolie	1.57	0.69	-0.24
Michalovce	1.67	0.28	-0.11
Rožňava	1.38	-0.74	0.25
Sobrance	1.65	-0.34	-1.03
Spišská Nová Ves	1.97	1.15	-0.72
Trebišov	1.28	1.76	-0.05

KORELAČNÁ MATICA

	UR	RO	RI	RUV	EO	LU13.48	LU48	D
UR	1	-0.84	0.02	0.58	0.04	0.67	0.81	0.01
RO	-0.84	1	0.02	-0.37	-0.09	-0.73	-0.68	-0.06
RI	0.02	0.02	1	0.02	0.03	-0.43	-0.11	-0.95
RUV	0.58	-0.37	0.02	1	-0.02	0.32	0.43	-0.03
EO	0.04	-0.09	0.03	-0.02	1	0.03	0.07	-0.03
LU13.48	0.67	-0.73	-0.43	0.32	0.03	1	0.54	0.48
LU48	0.81	-0.68	-0.11	0.43	0.07	0.54	1	0.07
D	0.01	-0.06	-0.95	-0.03	-0.03	0.48	0.07	1

Pre porovnanie uvádzam ortogonálnu rotáciu VARIMAX, a kosouhlú rotáciu COVARIMIN:

```
       *** Factor Analysis ***
Sums of squares of loadings:
  Factor1 Factor2   Factor3
 2.772751 2.13491 0.7772286

The number of variables is 8 and the number of observations is 78

Test of the hypothesis that 3 factors are sufficient
versus the alternative that more are required:
The chi square statistic is 10.42 on 7 degrees of freedom.
The p-value is 0.166

Component names:

 "loadings" "uniquenesses" "correlation" "criteria" "factors" "dof" "method"

 "center" "scale" "n.obs" "scores" "call"

Call:
factanal(x =  ~ UR + RO + RI + RUV + EO + LU13.48 + LU48 + D, factors = 3,
       method = "mle", data = DECEMBER00U, scores = T, type =
       "regression", rotation = "varimax", na.action = na.omit, control =
       list(iter.max = 20, unique.tol = 0.0001))

Importance of factors:
                 Factor1   Factor2    Factor3
   SS loadings 2.7727514 2.1349103 0.77722855
Proportion Var 0.3465939 0.2668638 0.09715357
Cumulative Var 0.3465939 0.6134577 0.71061129

The degrees of freedom for the model is 7.

Uniquenesses:
     UR         RO         RI       RUV        EO   LU13.48      LU48
 4e-010 0.02412681 0.08868414 0.6120188 0.9851462 0.2467416 0.3445094
          D
 0.01390979

Loadings:
        Factor1 Factor2 Factor3
     UR  0.971           0.240
     RO -0.691          -0.701
     RI         -0.949   0.106
    RUV  0.617
     EO                  0.119
LU13.48  0.596   0.491   0.396
   LU48  0.785           0.190
```

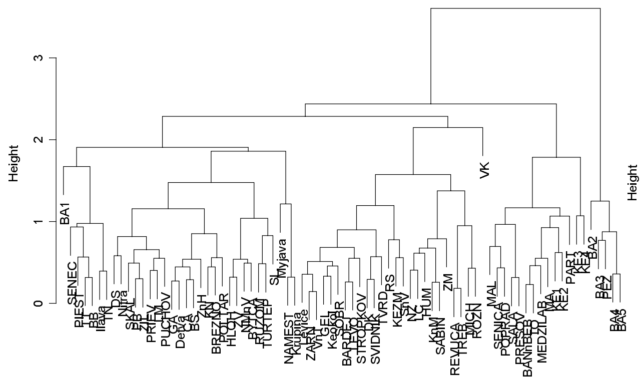
```
      D          0.990
```

***** Factor Analysis *****

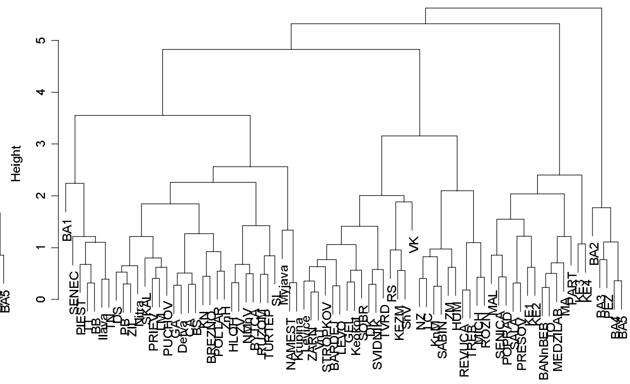
```
Sums of squares of loadings:
  Factor1  Factor2   Factor3
 2.560636 2.146653 0.5899978

The number of variables is 8 and the number of observations is 78

Test of the hypothesis that 3 factors are sufficient
versus the alternative that more are required:
The chi square statistic is 10.42 on 7 degrees of freedom.
The p-value is 0.166

Component names:

 "loadings" "uniquenesses" "correlation" "criteria" "factors" "dof" "method"

 "center" "scale" "n.obs" "scores" "call"

Call:
factanal(x =  ~ UR + RO + RI + RUV + EO + LU13.48 + LU48 + D, factors = 3,
       method = "mle", data = DECEMBER00U, scores = T, type =
       "regression", rotation = "covarimin", na.action = na.omit, control =
       list(iter.max = 20, unique.tol = 0.0001))

Importance of factors:
                   Factor1   Factor2    Factor3
     SS loadings 2.5606357 2.1466533 0.58999780
  Variable Index 0.3200795 0.2683317 0.07374972
Cumulative Index 0.3200795 0.5884111 0.66216085

The degrees of freedom for the model is 7.

Uniquenesses:
     UR         RO         RI       RUV        EO   LU13.48      LU48
 4e-010 0.02412681 0.08868414 0.6120188 0.9851462 0.2467416 0.3445094
         D
 0.01390979

Loadings:
        Factor1 Factor2 Factor3
     UR  0.963          -0.121
     RO -0.593           0.634
     RI         -0.951  -0.104
    RUV  0.651           0.151
     EO                 -0.122
LU13.48  0.507   0.490  -0.334
   LU48  0.772
      D          0.993

Component/Factor Correlations:
        Factor1 Factor2 Factor3
Factor1  1.000  -0.109   0.272
Factor2 -0.109   1.000   0.007
Factor3  0.272   0.007   1.000
```

Teraz nasledujú výsledky zhlukovania – dendrogramy pre rôzne metódy zhlukovania v uvedenom poradí: AVERAGE, COMPLETE, SINGLE, WEIGHTED, pri použití euklidovskej vzdialenosti.

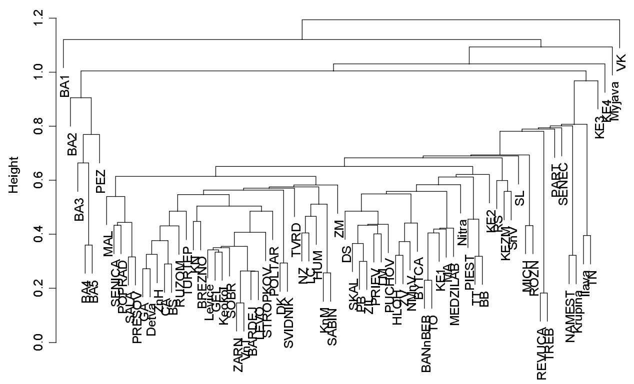
AVERAGE



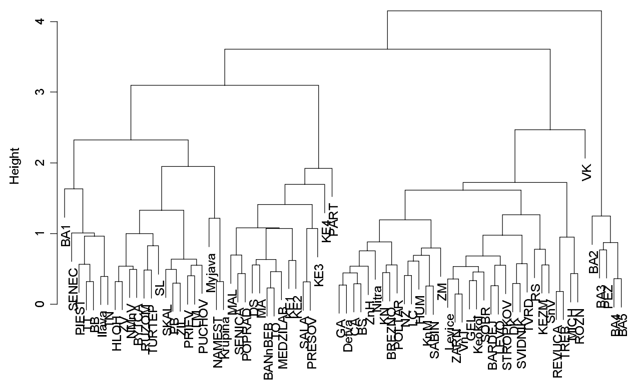
COMPLETE



SINGLE



WEIGHTED



Tabuľka príslušnosti do zhlukov:

Okres	Cluster
Bratislava I	3
Bratislava II	3
Bratislava III	3
Bratislava IV	3
Bratislava V	3
Malacky	1
Pezinok	3
Senec	3
Dunajská Streda	1
Galanta	4
Hlohovec	4
Pieštany	3
Senica	1
Skalica	4
Trnava	3
Bánovce n/Bebravou	1
Ilava	3
Myjava	4
Nové Mesto n/Váhom	4
Partizánske	1
Považská Bystrica	4
Prievidza	3
Púchov	3
Trencín	3
Komárno	1
Levice	4
Nitra	1
Nové Zámky	1
Šala	1
Topolcany	1
Zlaté Moravce	1
Bytca	4
Cadca	4
Dolný Kubín	4
Kysucké Nové Mesto	1
Liptovský Mikuláš	4
Martin	1
Námestovo	4
Ružomberok	4
Turcianske Teplice	4
Tvrdošín	4
Žilina	1
Banská Bystrica	3
Banská Štiavnica	4
Brezno	1
Detva	4
Krupina	4
Lucenec	1
Poltár	4
Revúca	1
Rimavská Sobota	2
Velký Krtíš	4
Zvolen	4
Žarnovica	4
Žiar nad Hronom	4
Bardejov	4
Humenné	1
Kežmarok	4
Levoca	4
Medzilaborce	1
Poprad	1
Prešov	1
Sabinov	1
Snina	2
Stará Lubovna	4
Stropkov	4
Svidník	4
Vranov nad Toplou	4
Gelnica	1
Košice I	1
Košice II	1
Košice III	1
Košice IV	1
Košice - okolie	1
Michalovce	2
Rožnava	1
Sobrance	4
Spišská Nová Ves	2
Trebišov	1