

Explorácia disociatívnych skúseností u ľudí z diagnózou depresie pomocou faktorovej analýzy a klasifikačných stromov

Miloslav Torda

8 júna 2017

Úvod

Mojim cieľom bolo vyskúšať si metódy, s ktorými som ešte nepracoval. Preto som si zvolil faktorovú analýzu a klasifikačné stromy.

Dáta

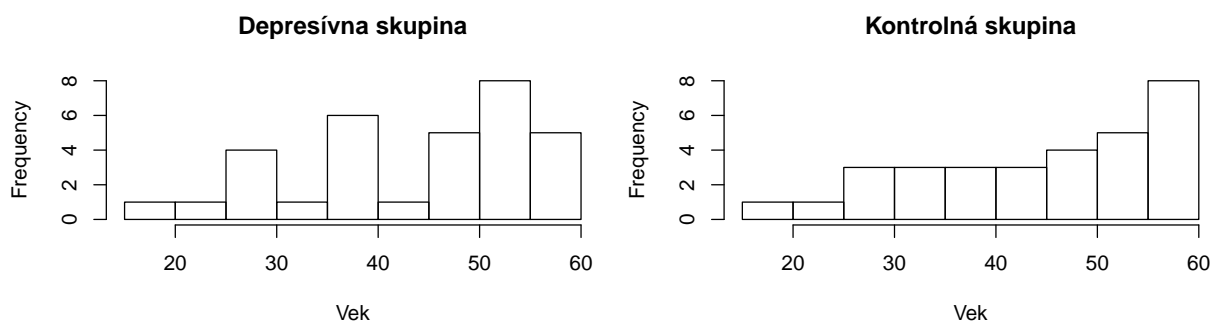
Dáta pochádzajú z mojej diplovej práce, ktorú som robil na psychológii. Názov práce je *Disociatívne procesy v symptomatológii depresii*. Robil som prieskum vzťahu depresívnych symptómov, symptómov post-traumatickej stresovej poruchy a symptómov disociačných porúch. Robil som to u hospitalizovaných ľudí s diagnózou depresie na psychiatrických oddeleniach v nemocniciach v Bratislave (Antolská, Mickiewiczova, Ružinov), Pezinku a Hronovciach.

Celá testová batéria pozostávala z:

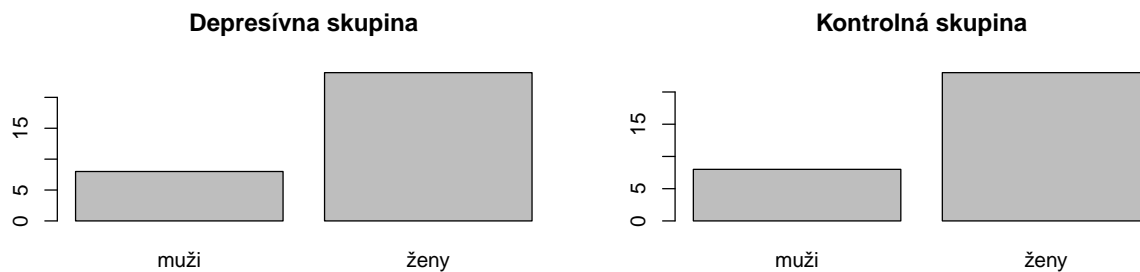
- Škála disociatívnych skúseností DES (Dissociative Experiences Scale)
- Posttraumatický zoznam PCL-C (PTSD checklist – civilian version)
- Beckova sebaopisová škála depresivity BDI-II (Beck Depression Inventory)
- Asociačný experiment
- Dotazník životných udalostí
- základné anamnestické otázky - vek, pohlavie, diagnóza, vzdelanie, partnerský status, zamestnanecký status a či má daný človek chronické telesné ochorenie.

Celková dátová vzorka pozostáva z 63 ľudí 32 z diagnózou depresie F32 a 31 bez psychiatrickej diagnózy.

Vekové rozloženie vzorky:



Rozloženie pohlavia vo vzorke:



Cieľ analýzy zvoleného dátového súboru

Z celého dátového súboru som si zvolil DES, čo je 28 položková sebaopisovacia škála, ktorá sa snaží merať ako často sa dané skúsenosti vyskytujú. Škála je nastavená tak, že pod každou otázkou je škála od 0% do 100% s delením každých 10% a ten čo vyplní škálu má zakrúžkovať kde sa asi nachádza s tým, že 0 je nikdy a 100 je stále.

Za cieľ som si stanovil urobiť exploráciu disociatívnych skúseností u ľudí s diagnózou depresie.

Opis aplikácie metód

Ako prvú som urobil faktorovú analýzu 28 otázok DES iba na vzorke ľudí z diagnózou depresie $N=32$. Následne som priradil, každú otázku ku jednému z faktorov na základe maxima z riadku matice loadingov jednoznačne pre každú otázku tj. jedna otázka iba ku jednému faktoru. Ako metódu na odhad matice loadingov som použil metódu *maximálnej vierohodnosti* a *varimaxovú* metódu rotácie matice loadingov.

Z toho som urobil agregáty otázok a hodnoty odpovedí respondentov som sčítal na základe príslušností ku danému faktoru. Toto si myslím, že je legitímne pretože vyhodnocovanie DES sa robí týmto spôsobom tj. sumáciou všetkých zložiek dotazníka. Skóre nad 20 sa považuje ako indikátor disociatívnej poruchy. V bežnej populácii dospelých je všeobecne pod 10.

V podstate som si vytvoril z dát na základe faktorovej analýzy nové features. A na tieto features som použil klasifikačné stromy. Keďže mám k dispozícii aj hodnoty DES aj u ľudí bez psychiatrickej diagnózy, použil som na klasifikáciu všetkých 63 respondentov a klasifikácia bola robená na základe toho či má alebo nemá diagnózu depresie a rozhodovania boli robené na základe vytvorených features. Pri klasifikácii 1 znamená, že daný človek má diagnózu depresie a 0, že nemá.

Výsledky

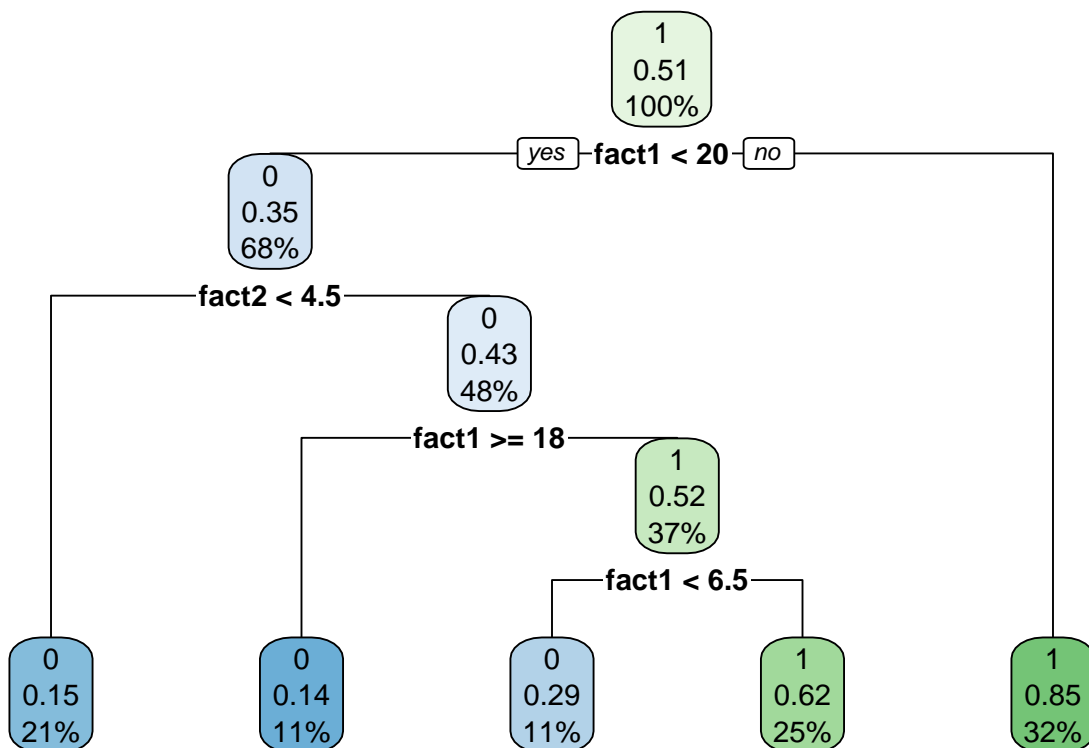
Najväčším problémom bolo zvoliť počet faktorov. Jeden spôsob bol postupne zvyšovať počet faktorov.

2 faktory

Zatriedenie jednotlivých položiek v prípade 2 faktorov:

```
## des01 des02 des03 des04 des05 des08 des09 des10 des16 des17 des21 des22
##      1      1      1      1      1      1      1      1      1      1      1      1
## des24 des25 des26 des27 des06 des07 des11 des12 des13 des14 des15 des18
##      1      1      1      1      2      2      2      2      2      2      2      2
## des19 des20 des23 des28
##      2      2      2      2
```

Klasifikačný strom



počet správnych klasifikácií

```
## [1] 49
```

3 faktory

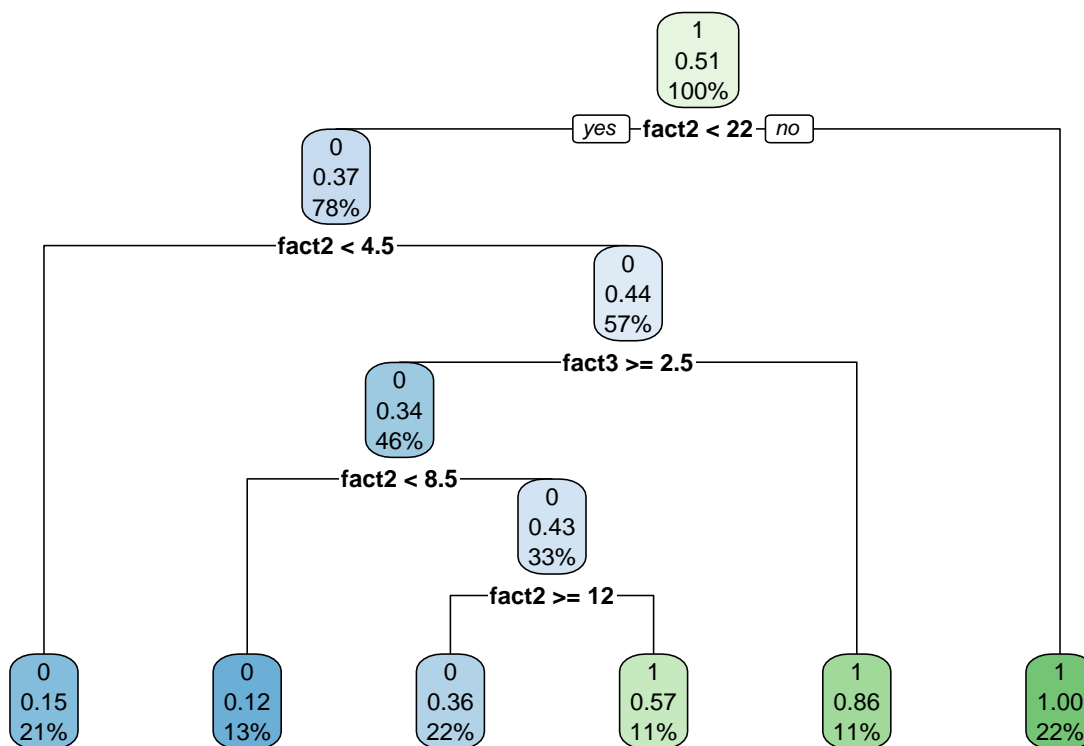
Zatriedenie jednotlivých položiek v prípade 3 faktorov:

```
## des01 des03 des04 des05 des08 des09 des10 des16 des21 des24 des25 des26
##      1      1      1      1      1      1      1      1      1      1      1      1
## des27 des02 des07 des11 des12 des13 des14 des15 des17 des18 des20 des22
##      1      2      2      2      2      2      2      2      2      2      2      2
## des28 des06 des19 des23
##      2      3      3      3
```

Zmena oproti 2 faktorovému modelu:

- faktor 1: -2,-17,-22,
- faktor 2: +2,-6,+17 -19,+22,-23
- faktor 3: +6, +19,+23

Klasifikačný strom



počet správnych klasifikácií

```
## [1] 51
```

4 faktory

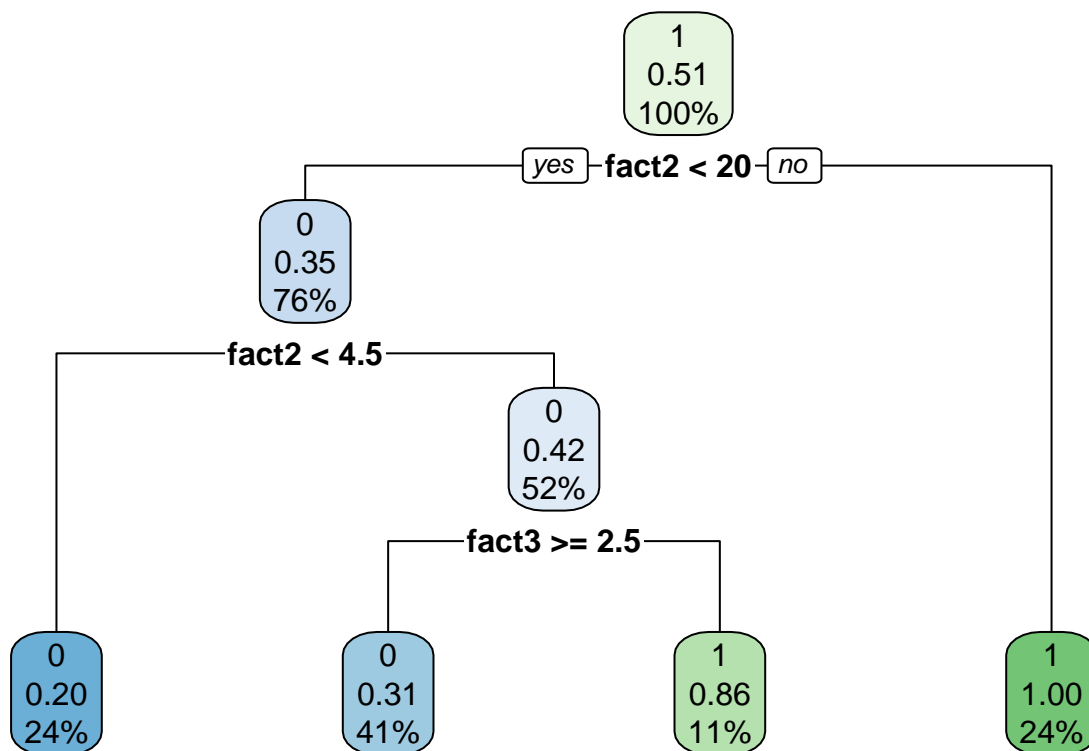
Zatriedenie jednotlivých položiek v prípade 4 faktorov:

```
## des01 des03 des04 des05 des08 des09 des10 des16 des21 des24 des25 des26
##      1      1      1      1      1      1      1      1      1      1      1      1
## des27 des02 des07 des11 des12 des13 des14 des15 des18 des20 des22 des28
##      1      2      2      2      2      2      2      2      2      2      2      2
## des06 des19 des23 des17
##      3      3      3      4
```

Zmena oproti 3 faktorovému modelu:

- faktor 1: 0
- faktor 2: -17
- faktor 3: 0
- faktor 4: +17

Klasifikačný strom



počet správnych klasifikácií

```
## [1] 51
```

5 faktorov

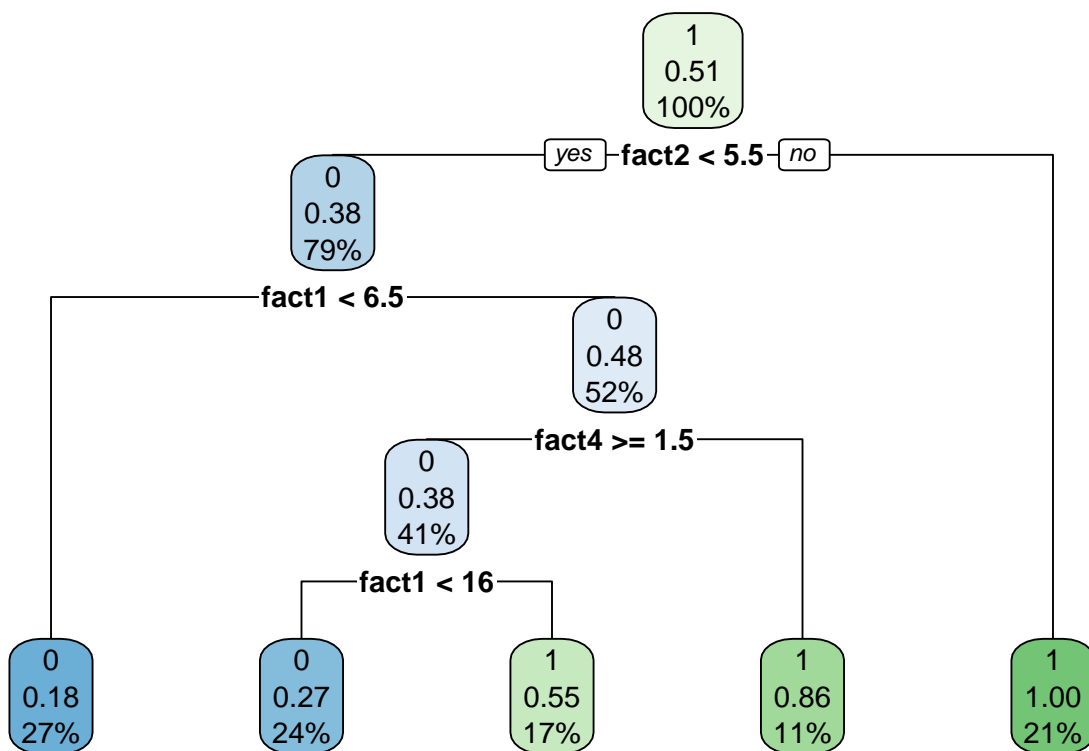
Zatriedenie jednotlivých položiek v prípade 5 faktorov:

```
## des01 des02 des03 des04 des05 des08 des09 des10 des16 des21 des24 des25
##      1      1      1      1      1      1      1      1      1      1      1      1
## des26 des27 des28 des07 des11 des13 des18 des20 des12 des14 des15 des22
##      1      1      1      2      2      2      2      2      3      3      3      3
## des06 des19 des23 des17
##      4      4      4      5
```

Zmena oproti 4 faktorovému modelu:

- faktor 1: +2
- faktor 2: -2,-12,-14,-15,22
- faktor 3: -6,+12,+14,+15,-19,+22,-23
- faktor 4: +6,-17,+19,+23
- faktor 5: +17

Klasifikačný strom



počet správnych klasifikácií

```
## [1] 50
```

6 faktorov

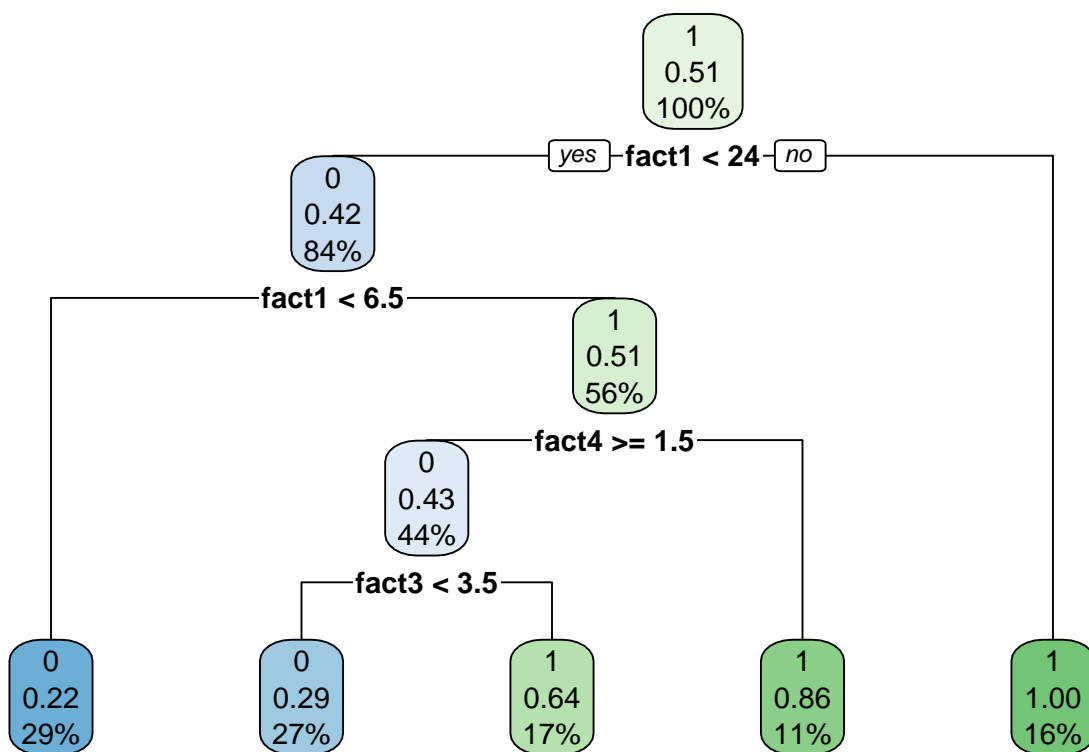
Zatriedenie jednotlivých položiek v prípade 6 faktorov:

```
## des01 des02 des03 des04 des05 des08 des09 des10 des16 des21 des24 des25
##      1      1      1      1      1      1      1      1      1      1      1      1
## des26 des27 des28 des12 des14 des15 des22 des11 des13 des18 des20 des06
##      1      1      1      2      2      2      2      3      3      3      3      4
## des19 des23 des07 des17
##      4      4      5      6
```

Zmena oproti 5 faktorovému modelu:

- faktor 1: 0
- faktor 2: faktor 3
- faktor 3: faktor 2, -7
- faktor 4: 0
- faktor 5: +7
- faktor 6: faktor 5

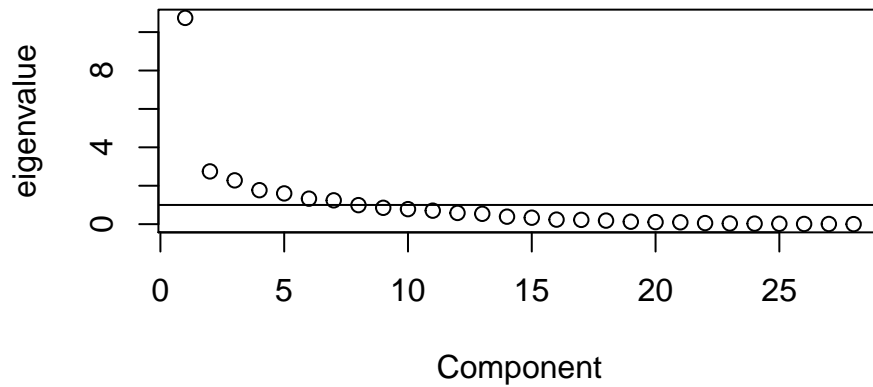
Klasifikačný strom



počet správnych klasifikácií

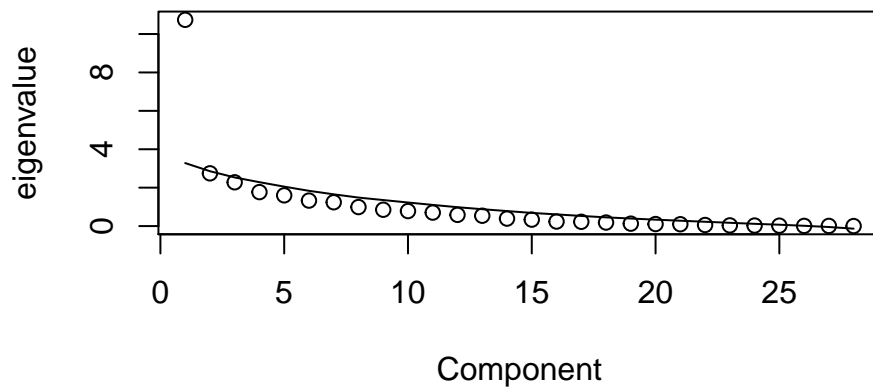
```
## [1] 49
```

Jedným zo spôsobov ako určiť počet faktorov je tzv. pravidlom K1. tj. zobrať ako počet faktorov počet vlastných čísel korelačnej matice väčších ako 1.



V tomto prípade by bol počet faktorov 7. Veľmi sa ale neodporúča toto pravidlo používať.

Modifikáciou pravidla K1 je tzv. paralelná analýza. Vytvorí sa náhodná matica rozmeru datasetu s tým, že rozdelenie jednotlivých prvkov matice je $N(0,1)$. Z toho sa zráta korelačná matica a z nej sa vyrátajú vlastné čísla. Toto sa zopakuje n krát a vyráta sa potom polohová štatistika (priemer, median etc) vlastných čísel simulovaných korelačných matíc. Potom počet faktorov je počet vlastných čísel väčších ako príslušnej polohovej štatistiky vlastného čísla zo simulovaných korelačných matíc.



Ak by sme zvolili ako polohovú štatistiku priemer tak počet faktorov by mal byť

```
## [1] 3
```


Diskusia

Už bola robená faktorová analýza DES. Na populácií ľudí zneužívajúcich alkohol a ine omamné látky zistili 3 faktory:

1. amnestická disociácia: otázky - 3,4,5,6,8,10,25,26
2. absorpcia/imaginatívne pohrúženie: otázky - 2,14,15,16,17,18,20,22,23
3. depersonalizácia/derealizácia: otázky - 7,11,12,13,27,28

Z časti o výsledkoch vidno, že otázky, ktoré boli pridané ako agregát s názvom *fact1* sa zvyšovaním počtu faktorov moc nemení. Najväčšia zmena je pri prechode na troj-faktorový model. Taktiež pri klasifikačných stromoch moc nevystupuje ako rozhodovací parameter. Akurát pri 2-faktorovom modeli. Potom, už nie. Pravdepodobne tým, že pri prechode z 2-faktorového modelu na prešli z faktora 1 do faktora 2 otázky 2,17,22.

Zaujímavejším je faktor 2. Už pri 3 faktorovom modeli sa javí tento faktor ako dobrý klasifikátor pre istú skupinu ľudí s depresiou. Toto pozorovanie pretrváva pri zvyšovaní počtu faktorov na 4 a 5 faktorový model. Pri 6 faktorovom modeli sa otázky z faktora 2 presunuli do faktora 3 a všeobecne sa to premiešalo. Osobne by som sa zastavil pri 4 faktorovom modeli kvôli tomu, že najviac faktor 2 pri tomto modeli oddeľuje najväčší počet ľudí s vysokým skóre pri tomto faktore a pretože faktor 4 je iba jedna otázka, ktorá nemá vplyv na klasifikáciu pri klasifikačnom strome. Čo znamená, že to je v podstate 3 faktorový model, ktorý vychádza ako dobrý počet faktorov z paralelnej analýzy.

Pri 3,4 faktorovom modeli klasifikačný strom ukazuje jednu skupinu ľudí s diagnózou depresie, ktorá sa oddeľuje rovno s koreňového uzla. Špecifické pre túto skupinu je vysoké skórovanie vo faktore 2. (nad 20. pripomeniem, že všeobecné skóre v DES nad 10 naznačuje disociatívnu poruchu). Podobná situácia bola aj pri 2 faktorovom modeli ,akurát že v tomto prípade to bol faktor 1 a pri tejto klasifikácii spadli do klasifikačnej množiny aj ľudia bez diagnózy depresie. Hlavný rozdiel oproti 2 faktorovému modelu a 3 faktorovému modelu vidím v tom, že otázky 2, 17 a 22 prešli do faktora 2. Pri ďalšom zvyšovaní faktorov prešla otázka 17 do faktora číslo 4. Teda akoby vysoké skóre v otázkach 2 a 22 bolo špecifické pre istú podskupinu depresívnych pacientov.

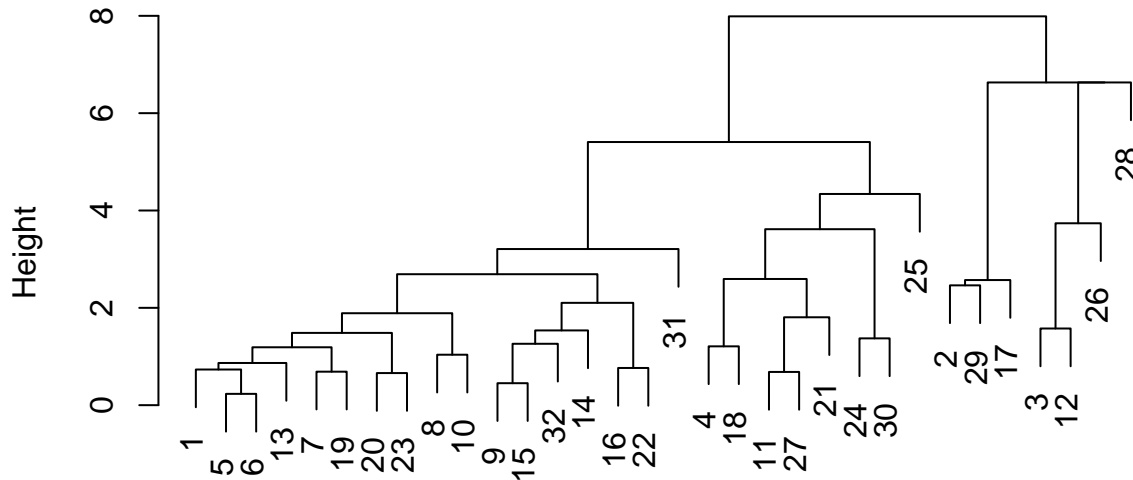
Rozdelenie otázok do faktorov pre 4 faktorový model:

1. faktor - 1,3,4,5,8,9,10,16,21,24,25,26,27
2. faktor - 2,7,11,12,13,14,15,18,20,22,28
3. faktor - 6,19,23
4. faktor - 17

Z vyššie uvedeného máme, že faktor 2 pozostáva z kombinácie otázok na skúsenosti s absorpcie/ imaginatívneho pohrúženia a depersonalizácie a derealizácie. Faktor 1 pozostáva väčšinou z otázok z amnestickej disociácie.

Napadlo mi, či sa neukáže táto skupina s vysokým skórovaním vo faktore 2 ako nejak odlišná od ostatných ľudí s diagnózou depresie a či sa to neukáže pri hierarchyckom klustrovaní. Uvedomujem si, že to nie je úplne vhodná metóda na tieto dáta, ale aj tak som použil divizívne hierarchycke klustrovanie na faktoroch zo 4 faktorového modelu a skupine ľudí s diagnózou depresie.

Dendrogram



cum_fact[1:32,]
Divisive Coefficient = 0.81

Na dendrograme vidno hned prvý zhluk, ktorý algoritmus oddelil ako najmenej podobný voči ostatným objektom. (najviac v pravo). Keď sa pozrieme na hodnoty faktora 2 u týchto ľudí tak si môžeme všimnúť, že špecifikom u týchto ľudí je hodnota faktora 2, ktorá je nad 26.

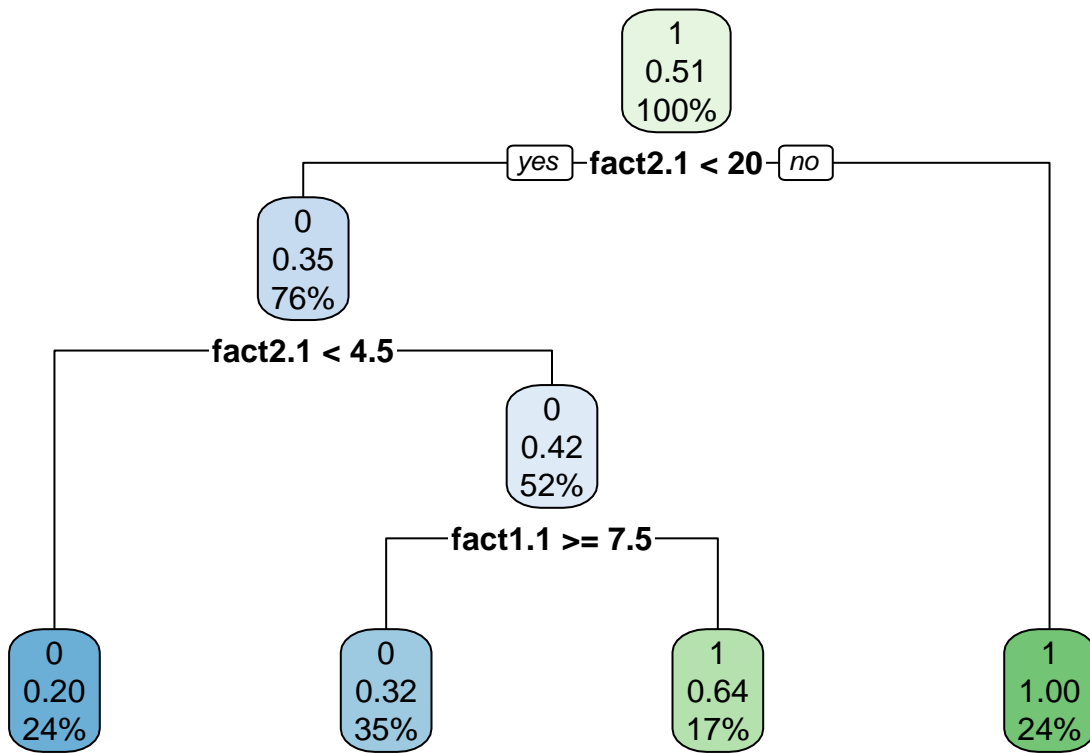
Je dosť možné, že analýza odhalila podskupinu ľudí z diagnózou depresie, u ktorej je vysoké skóre v otázkach z faktora 2. tj. absorpcie/ imaginatívneho pohrúženia a depersonalizácie a derealizácie. Uvažuje sa o súvisi post traumatickej stresovej poruchy a depresívnych symptómov. A práve u ľudí, ktorí prešli silným traumatickým zážitkom sa objavujú disociačné symptómy ako forma obranného mechanizmu. Taktiež existuje skupina ľudí ktorý majú diagnózu depresie, ale sú dlhodobo rezistentný na antidepressíva a objavujú sa u nich v minulosti traumatické zážitky.

Ešte poznámka ku klasifikáciam. Disociatívne skúsenosti sú v istej miere bežné všeobecne v populácii. Preto nebol schopný algoritmus klasifikačných stromov presne klasifikovať. Inými slovami pozorované objekty neboli lineárne separovateľné.

Na záver som ešte urobil 2 faktorovú analýzu na otázkach z faktorov 1 a 2 z 4 faktorového modelu

```
## des01 des03 des04 des05 des08 des09 des10 des16 des21 des24 des25 des26
##      1      1      1      1      1      1      1      1      1      1      1      1
## des27 des02 des07 des11 des12 des13 des14 des15 des18 des20 des22 des28
##      1      2      2      2      2      2      2      2      2      2      2      2
```

a otázky ostali zaradené v rovnakých faktoroch ako pri 4 faktorovom modeli. Podobne na tom bol aj klasifikačný strom



Appendix

K-means

