# Láncfűrészek paramétereinek összehasonlítása a többváltozós statisztika módszereivel

**Horváth-Szováti Erika**
NymE EMK Matematikai Intézet
hsze@emk.nyme.hu

*Összefoglaló.* A főkomponens-analízis és faktoranalízis sok esetben hatékonyan használható a kísérletek kiértékelésében. A két módszer nagyon hasonlít egymásra, de van közöttük néhány fontos különbség. A két eljárást egy műszaki területről vett példa segítségével szeretnénk bemutatni.

*Abstract.* The principal component analysis and factor analysis can be used effectively in many cases, evaluation of the experiments. The two methods are very similar, but there are some important differences between them. We want these two methods with an example in a technical area to illustrate.

## 1. Bevezetés

A többváltozós statisztika módszerei közül a főkomponens-analízis (*Principal component analysis; PCA*) és faktoranalízis *(Factor analysis; FA)* alkalmazására mutatunk egy példát. Ezeknek az eljárásoknak a használata akkor javasolt, ha nagyszámú független változóval dolgozunk, mert ilyenkor elvész az ábrázolhatóság előnye, és lecsökken a változók tényleges függetlenségének valószínűsége. A független változók között kölcsönös, többirányú összefüggések lehetnek, azaz multikollinearitás állhat fenn. Mind a *PCA*, mind a *FA* esetén ugyanaz a cél: dimenziószám csökkentés a legkisebb információvesztés mellett. Ezt úgy érhetjük el, hogy kevesebb számú új változót vezetünk be (közben a varianciát maximalizáljuk). Az új változókat *PCA* esetén nem mindig lehetséges azonosítani, *FA* alkalmazásakor viszont elvárás, hogy megnevezzük őket. Ezt követően az új változók (főkomponensek/faktorok) alkotta új bázisban a régi változók és a mérési eredmények koordinátáit felírjuk, ebből próbálunk következtetéseket levonni, illetve így az adathalmaz ábrázolása kényelmesebb és sokkal több információt hordoz. A példára vonatkozó számításokat a *STATISTICA 11* programcsomag segítségével végeztük.

Mind a *PCA*, mind a *FA* kiindulhat a kovariancia-, illetve korrelációs mátrix elemzéséből. Mindkét esetben a kovariancia-/korrelációs-mátrix sajátértékei alapján határozzuk meg a főkomponensek, illetve faktorok számát. A *PCA* és *FA* módszerei között lényeges eltérések is vannak. A főkomponensek az eredeti változók olyan lineáris kombinációi, amelyek minél nagyobb számú eredeti változóval állnak szoros korrelációban. Mivel legtöbbször az eredeti változók is nagyon sokfélék, így a lineáris kombinációik csak ritkán értelmezhetők. A faktoranalízis az adathalmaz mögött meghúzódó lineáris háttér-összefüggéseket tételez fel, és a faktorok az új „háttérváltozók". A faktorokat mindenképpen értelmeznünk kell, ebben segítenek a különböző rotációs eljárások. Ilyenkor a háttérváltozók koordináta-rendszerét addig forgatjuk, amíg olyan helyzetbe nem kerül, hogy a mérési eredmények koordinátái csak egy-egy tengely (faktor) irányában rendelkeznek magas koordinátákkal, így a faktorok

értelmezhetővé válnak. A *STATISTICA 11* programcsomag több különböző típusú faktorrotációt tesz lehetővé. A legismertebb derékszögű faktorrotációs eljárások a *varimax*, *quartimax* és *equamax* forgatás, a ferdeszögű forgatások közül leggyakoribbak a *direct oblimin* és *promax* forgatások. Legtöbbször a varimax rotációt használjuk, ez szinte mindig célravezető. A faktorok számának viszonylag szubjektív meghatározásából, illetve a választható sokféle forgatásból adódóan a faktoranalízis modelljének nagyon nagyszámú alternatív megoldása van.

## 2. Láncfűrészek műszaki paramétereinek és árának összehasonlítása PCA és FA segítségével

*Husqvarna* benzines láncfűrészek műszaki paramétereit és árát hasonlítjuk össze főkomponens-analízis és faktoranalízis segítségével. A következő változókat vettük be a vizsgálatba:

Var1: lökettérfogat ($cm^3$);
Var2: teljesítmény (kW);
Var3: teljesítmény (LE);
Var4: a motor fordulatszáma alapjáraton (fordulat/perc);
Var5: a motor maximális fordulatszáma (fordulat/perc);
Var6: láncsebesség maximális teljesítményen (m/s);
Var7: egyenértékű vibrációs szint (ahv, eq) elülső fogantyú ($m/s^2$);
Var8: egyenértékű vibrációs szint (ahv, eq) hátsó fogantyú ($m/s^2$);
Var9: zajszint (dB(A));
Var10: hangteljesítményszint (LWA; dB(A));
Var11: tömeg (kg);
Var12: ár (Ft).

Mind a *PCA*, mind a *FA* alkalmazhatósága ellenőrizhető a korrelációs mátrix értékeinek vizsgálatával. Kívánatos, hogy minél több korreláció abszolút értéke legyen magasabb, mint 0,3. Ezt a 12x12-es mátrixot itt nem közöljük, csupán az észrevételt, hogy szinte csak a Var5 és Var9 változók sorában (oszlopában) vannak az előbbi kritériumnak nem megfelelő elemek, a többi érték a változók között erős, vagy közepesen erős lineáris korrelációs kapcsolatot mutat.

### 2.1. Az adatok elemzése PCA-val

A korrelációs mátrix sajátértékei és a kumulatív variancia segítségével dönthető el, hogy hány főkomponenssel dolgozunk tovább (1. táblázat). Ebben az 1-nél nagyobb sajátértékek száma, illetve a minimum kb. 80%-os kumulatív variancia a meghatározó. Ezek alapján esetünkben 2 főkomponenst érdemes választani, a két főkomponens együtt a teljes adathalmaz varianciájának 85%-át magyarázza. A változók és a főkomponensek közötti korreláció értékeit (vagyis a faktorsúlyokat) a 2. táblázat mutatja. Ez alapján azt mondhatjuk, hogy a vizsgált változók két csoportba (főkomponensbe) sorolhatók. Az első főkomponenssel nagyon szoros negatív korrelációt mutatnak a Var1, Var2, Var3, Var11, Var12 változók, szoros negatív korrelációt a Var6, Var7, Var8 változók, és az első főkomponens erős pozitív korrelációban áll a Var4 változóval. A második főkomponens a Var5 és Var9 változókkal van erős negatív korrelációban.

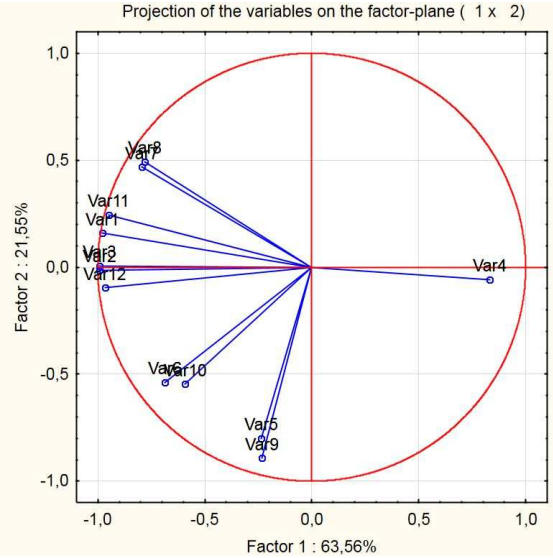
| Eigenvalues of correlation matrix, and related statistics<br>Active variables only | | | | |
|---|---|---|---|---|
| Value number | **Eigenvalue** | % Total variance | Cumulative Eigenvalue | Cumulative % |
| **1** | 7,627127 | 63,55939 | 7,62713 | 63,5594 |
| 2 | 2,585850 | 21,54875 | 10,21298 | 85,1081 |
| 3 | 0,880319 | 7,33600 | 11,09330 | 92,4441 |
| 4 | 0,424628 | 3,53857 | 11,51792 | 95,9827 |
| 5 | 0,287477 | 2,39564 | 11,80540 | 98,3783 |
| 6 | 0,088861 | 0,74051 | 11,89426 | 99,1188 |
| 7 | 0,044033 | 0,36695 | 11,93830 | 99,4858 |
| 8 | 0,034643 | 0,28869 | 11,97294 | 99,7745 |
| 9 | 0,012806 | 0,10672 | 11,98574 | 99,8812 |
| 10 | 0,011746 | 0,09788 | 11,99749 | 99,9791 |
| 11 | 0,002432 | 0,02027 | 11,99992 | 99,9993 |
| 12 | 0,000079 | 0,00066 | 12,00000 | 100,0000 |

**1. táblázat. A korrelációs mátrix sajátvektorainak sajátértékei**

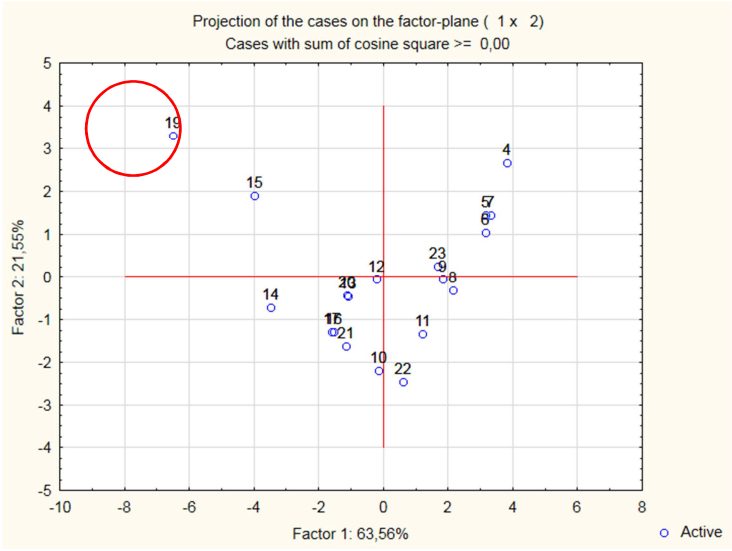
A főkomponensek elnevezésével a további elemzés egyszerűbb lenne, de ez azonban nem feltétlenül szükséges. Az elnevezés akkor könnyebb, ha a faktorok csupán egy-két változóval mutatnak szorosabb (akár negatív, akár pozitív) kapcsolatot. Esetünkben a főkomponensek értelmezése nem ilyen egyértelmű, az első főkomponenst esetleg „negatív robosztusság” főkomponensnek, a másodikat „fülkímélőség mértéke” főkomponensnek nevezhetjük. A változók faktor-koordinátáit egységkörön ábrázolva a 1. ábrán láthatjuk. A változók közötti lineáris korreláció mértéke a vektorok által közbezárt szög koszinuszával arányos. Például a Var5-Var9, Var6-Var10, Var1-Var11, Var 2-Var3-Var12, Var7-Var8 változók vektorai kicsi szöget zárnak be, közöttük nagyon szoros a pozitív korreláció. Amely vektorok között $90^o$ a közbezárt szög, azok a változók korrelálatlanok, de nem feltétlenül függetlenek. Ez azt jelenti, hogy nincs közöttük lineáris kapcsolat, viszont egyéb függvénykapcsolat lehetséges. Esetünkben például a Var1 és Var11 vektorok közelítőleg merőlegesek a Var5 és Var9 vektorokra, ez alapján mondhatjuk, hogy ezek a változók korrelálatlanok egymással. A Var2-Var3 és a Var4 változók vektorai közelítőleg $180^o$-os szöget zárnak be egymással, közöttük a korreláció értéke –1. Az egyes mérések faktor-koordinátái az 2. ábrán láthatjuk. Ez sok információt hordoz, pl. a legrobosztusabb és egyben az összes közül „legfülkímélőbb” fűrész a 19. sorszámú, a legzajosabb a 22., amely egy közepesen robusztus fűrész, a 4. sorszámú pedig csak egy kis hobbifűrész, és minimális az általa kibocsátott zajterhelés, és így tovább.

| | Factor coordinates of the variables | |
|---|---|---|
| Variable | **Factor 1** | Factor 2 |
| **Var1** | -0,978678 | 0,159274 |
| Var2 | -0,991602 | -0,013815 |
| Var3 | -0,991131 | 0,005830 |
| Var4 | 0,830783 | -0,057137 |
| Var5 | -0,237294 | -0,799889 |
| Var6 | -0,687961 | -0,537629 |
| Var7 | -0,795482 | 0,467865 |
| Var8 | -0,780855 | 0,494073 |
| Var9 | -0,233723 | -0,892897 |
| Var10 | -0,592591 | -0,546530 |
| Var11 | -0,949116 | 0,244617 |
| Var12 | -0,966831 | -0,096539 |

**2. táblázat. A változók és a faktorok közötti korreláció (faktorsúlyok)**




**1. ábra. A változók faktor-koordinátái egységkörön ábrázolva**




**2. ábra. A változók faktor-koordinátái a faktorok síkjában ábrázolva**

## 2.2. Az adatok elemzése FA-val

A 2. táblázatban lévő sajátértékeket alapján 2 faktort érdemes választani (2 db egynél nagyobb sajátérték van). A programcsomag által felkínált *scree-test* grafikont is meg szoktuk vizsgálni (itt hely hiányában nem közöljük), amelynek a „könyöke” a 3-as sajátértéknél látható, így ez alapján 3 faktort választanánk. A döntést az alapján hozzuk meg, hogy a vizsgálat során kapott 2 vagy 3 faktor értelmezhető jobban. A vizsgált adathalmaz esetében az elemzés lépéseinek többféle variációját (2, illetve 3 faktor, különböző forgatások) kipróbáltunk. A faktorok értelmezhetősége, illetve a fizikai (gépészeti) ismereteink alapján a kétfaktoros modell és a varimax rotáció mellett döntöttünk. A kapott eredmény az 5. táblázatban látható. Az első faktor a Var1-Var2-Var3, Var7-Var8, Var11-Var12 (lökettérfogat, kétféle teljesítmény és kétféle vibrációs adat, valamint a tömeg és az ár) változókkal nagyon erős pozitív, a Var4 (fordulatszám alapjáraton) változóval nagyon erős negatív korrelációban van. A második faktor a Var5-Var6 és Var9-Var10 (maximális fordulatszám, maximális láncsebesség, valamint zajszint és hangteljesítményszint)

változókkal viszonylag szoros pozitív korrelációban áll. (Megjegyezzük, hogy a második faktor oszlopában a Var10-nél lévő 0,6976 érték majdnem eléri a többi pirossal kiemelt, 0,7-nél nagyobb értéket, így ezt is figyelembe vettük.) Ezek alapján a következő módon nevezzük el a faktorokat: 1. faktor: teljesítmény-ár faktor, 2. faktor: zajterhelési faktor. Az egyes motorfűrészeknek ebben a kétfaktoros rendszerben lévő koordinátáit a 6. táblázat mutatja. Ebből leolvasható, hogy a teljesítmény-ár faktor (1. faktor) tekintetében legnagyobb koordinátával a 19. sorszámú, legkisebbel a 6. sorszámú motorfűrész rendelkezik. A 2. faktor-koordináták (zajterhelést mutató faktor) közül legnagyobb a 22. motorfűrész koordinátája, legkisebb pedig a 4. sorszámúé. Tehát a *FA* során kapott eredmény összhangban van a *PCA*-val kapott eredménnyel.

Factor Loadings (Varimax raw) (motorfűrészek)
Extraction: Principal components
(Marked loadings are >,700000)

| Variable | Factor 1 | Factor 2 |
|---|---|---|
| **Var1** | 0,981902 | 0,138017 |
| Var2 | 0,942932 | 0,307156 |
| Var3 | 0,948306 | 0,288255 |
| Var4 | -0,810376 | -0,191718 |
| Var5 | -0,010502 | 0,834279 |
| Var6 | 0,497655 | 0,717409 |
| Var7 | 0,898422 | -0,211011 |
| Var8 | 0,892223 | -0,240377 |
| Var9 | -0,041485 | 0,922047 |
| Var10 | 0,403933 | 0,697637 |
| Var11 | 0,978968 | 0,047747 |
| Var12 | 0,894751 | 0,378818 |
| Expl.Var | 7,184080 | 3,028896 |
| Prp.Totl | 0,598673 | 0,252408 |

**3. táblázat. Faktorsúlyok kétfaktoros modell esetén, varimax rotáció**

Factor Scores (motorfűrészek)
Rotation: Varimax raw
Extraction: Principal components

| Case | Factor 1 | Factor 2 |
|---|---|---|
| **4** | -0,821657 | **-1,99525** |
| 5 | -0,819273 | -1,20563 |
| 6 | **-0,900248** | -0,96009 |
| 7 | -0,881303 | -1,21540 |
| 8 | -0,798061 | -0,04291 |
| 9 | -0,641048 | -0,17487 |
| 10 | -0,354525 | 1,31215 |
| 11 | -0,660770 | 0,65486 |
| 12 | 0,070805 | 0,04880 |
| 13 | 0,296651 | 0,38089 |
| 14 | 1,072381 | 0,80045 |
| 15 | 1,734104 | -0,70547 |
| 16 | 0,301177 | 0,92195 |
| 17 | 0,318815 | 0,93370 |
| 19 | **2,860845** | -1,26529 |
| 20 | 0,312018 | 0,37123 |
| 21 | 0,103923 | 1,08214 |
| 22 | -0,662828 | **1,39699** |
| 23 | -0,531005 | -0,33824 |

**4. táblázat. Faktorsúlyok kétfaktoros modell esetén, varimax rotáció**

Mindenképpen szólnunk kell arról, hogy a faktoranalízis megbízhatósága a reziduális korrelációs mátrixszal vizsgálható, amely a mért adatok közötti, és a felállított modell segítségével számított értékek közötti korrelációkat hasonlítja össze. Ha a mátrixban a főátlón kívül nincs túl sok 0,1-nél lényegesen nagyobb érték (ezeket pirossal jelöli a program), akkor megállapítható, hogy a modell jól reprodukálja a mérési eredményeket. Példánkban a reziduális korrelációs mátrixot kielégítőnek találjuk (ennek közlése itt a terjedelmi korlátok miatt nem lehetséges), így a felállított modell megbízhatónak mondható.

## 3. Összefoglalás

Napjainkban legtöbb tudományterületen szinte elképzelhetetlenek a nívósabb publikációk statisztikai alkalmazások nélkül. A főkomponens-analízis és faktoranalízis a többváltozós adathalmazok vizsgálatában nagy segítséget nyújthat. Az eljárások matematikai hátterének vázlatos megismerése – különös tekintettel a közöttük lévő hasonlóságokra és különbségekre nagyon fontos, ezek hiányában a programcsomag felhasználója nehezen tudja értelmezni a kapott eredményeket és mindvégig bizonytalanságot fog érezni. Szeretnénk, ha a matematikai ismeretek szükségességét a szakmai tárgyakat oktató kollégáink közül is egyre többen felismernék, és ezáltal már az alapszintű matematika tananyagtól kezdve alaposabb tanulásra buzdítanák hallgatóinkat.

## Irodalomjegyzék

[1] **Fazekas, I.** (szerk.), Bevezetés a matematikai statisztikába, egyetemi jegyzet, Kossuth Egyetemi Kiadó, Debrecen (1997).
[2] **Füstös, L., Meszéna, Gy., Simonné Mosolygó, N.,** A sokváltozós adatelemzés matematikai módszerei, Akadémiai Kiadó, Budapest (1986).
[3] **Münnich, Á., Nagy, Á., Abari, K.,** Többváltozós statisztika pszichológus hallgatók számára. Bölcsész Konzorcium, Debrecen (2006). (http://psycho.unideb.hu/statisztika ISBN 9639704040.)
[4] STATISTICA 11, STATISTICA statisztikai adatelemző, analitikai szoftvercsalád, StatSoft.
[5] STATISTICA 11 software HELP.
[6] **Sváb, J.,** Többváltozós módszerek a biometriában. Mezőgazdasági Kiadó, Budapest (1979).
[7] **Szűcs, I.** (szerk.), Alkalmazott statisztika. Agroinform Kiadó, Budapest (2002).
[8] **Jahn, W. – Vahle, H.,** A faktoranalízis és alkalmazása. Közgazdasági és Jogi Könyvkiadó (1974).