Matematičke metode u marketingu. Eksploratorna faktorska analiza (EFA[1])

Lavoslav Čaklović
PMF-MO

2016

[1] Exploratory Factor Analysis

TEORIJSKA RAZRADA MODELA

Faktorska analiza donosi zaključke o latentnim (skrivenim, neuočljivim) varijablama kao što su: inteligencija, muzikalnost, navike potrošača... koje se ne mogu direktno mjeriti. One se mjere putem njihovih manifestacija u odabranom kontekstu.

$$
\begin{array}{rcccll}
\text{subjekti} & \downarrow & X_1 & \cdots & X_p \leftarrow & \text{manifestacije mjerljivo} \\
& s_1 & x_{11} & \cdots & x_{1p} & \\
& s_2 & x_{21} & \cdots & x_{2p} & \\
& \vdots & \vdots & \ldots & \vdots & \\
& s_m & x_{m1} & \cdots & x_{mp} &
\end{array}
$$

FA opisuje korelaciju među manifestinim varijablama $X_1, \ldots X_p$ u terminima manjeg broja latentnih **faktora** koji su uzrok varijabilnosti u podacima..

Primjer

Marketinška firma želi ustanoviti što utječe na vjernost kupaca 'svojim' dućanima. U tu svrhi kupci su ispunjavali anketu (80 pitanja). Menadžeri pretpostavljaju da ta vjernost počiva na nekoliko faktora: ljubaznosti osoblja, kvaliteti usluga, atmosferi u dućanu, širini ponude, kvaliteti proizvoda i njihovoj cijeni.

Faktorska analiza daje odgovor je li moguće svih 80 atributa svrstati u 6 grupa koje će biti prepoznate kao gore navedeni faktori.

ORTOGONALNI MODEL

$$X_1 - \mu_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \cdots \lambda_{1m}F_m + \epsilon_1$$
$$X_2 - \mu_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \cdots \lambda_{2m}F_m + \epsilon_2$$
$$\vdots$$
$$X_p - \mu_p = \lambda_{p1}F_1 + \lambda_{p2}F_2 + \cdots \lambda_{pm}F_m + \epsilon_p$$

λ_{ij} – **factor loading** i-tog atributa na j-tom faktoru

ϵ_i – nerastavljivi faktor (unique, specific)

ili u matričnom obliku

$$\mathbb{R}^r \ni (X - \mu) = \underset{\substack{\uparrow \\ \text{matrica punjenja (loadings matrix)}}}{\Lambda} \cdot F + \epsilon. \tag{1}$$

Pretpostavke ortogonalnog modela:

- $E(F) = 0, \quad Var(F) = E(FF^\tau) = I$ (2)

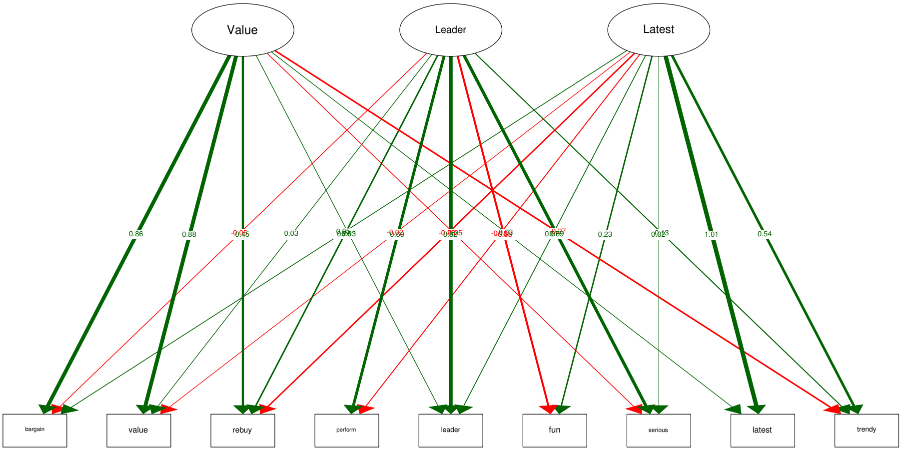
$\uparrow$

Reskaliranjem faktora se može svesti na taj slučaj tako da to nije ograničenje

- $E(\epsilon) = 0, \quad Var(\epsilon) = E(\epsilon\epsilon^\tau) = \Psi = diag(\psi_1, \dots, \psi_p)$ (3)
- F i ϵ su nekorelirane, tj. $Cov(F, \epsilon) = 0$ (4)
- Varijable ϵ_i nisu korelirane. (5)

Napomena. Nekoreliranost (ortogonalnost) faktora je zahtjev nepostojanja sinergije među njima (teško uhvatljiv pojam).

Dijagram


Value
Leader
Latest
0.86
0.88
0.03
0.23
1.01
0.54
bargain
value
rebuy
perform
leader
fun
serious
latest
trendy

Posljedice:

$$\begin{aligned} Cov(X,F) &= E((X-\mu)F^\tau) = E(\Lambda F F^\tau + \epsilon F^\tau) \\ &= \Lambda E(FF^\tau) + E(\epsilon F^\tau) = \Lambda \end{aligned} \tag{6}$$

odnosno[8]

$$\begin{aligned} \Sigma &:= E((X-\mu)(X-\mu)^\tau) \\ &= \Lambda E(FF^\tau)\Lambda^\tau + E(\epsilon F^\tau)\Lambda^\tau + \Lambda E(F\epsilon^\tau) + E(\epsilon\epsilon^\tau) \\ &= \Lambda\Lambda^\tau + \Psi. \end{aligned} \tag{7}$$

Specijalno, zbog dijagonalnosti Ψ,

$$\sigma_i^2 := Var(X_i) = \underbrace{\lambda_{i1}^2 + \lambda_{i2}^2 + \cdots + \lambda_{im}^2}_{\substack{\text{faktorsko objašnjenje varijance} \\ \text{communalities (oznaka: } h_i^2) \\ \text{što veće}}} + \underbrace{\psi_i}_{\substack{\text{specifična varijanca} \\ \text{uniqueness (oznaka: } u_i^2) \\ \text{što manja}}}$$

[8] Minimal residual FA.

Upute za korištenje

- broj faktora mali u odnosu na broj varijabli
- $\frac{p(p+1)}{2}$ varijanci i kovarijanci iz Σ su reproducirani preko $pm + p$ faktora. Jedinstvenost faktora je upitna.
- nije istina da je jednadžba $\Sigma = \Lambda\Lambda^\tau + \Psi$ rješiva za svako Σ. Napr. za $m = 1$ (jednofaktorski model) i

$$\Sigma = \begin{bmatrix} 1 & 0.9 & 0.7 \\ 0.9 & 1 & 0.4 \\ 0.7 & 0.4 & 1 \end{bmatrix}$$

  to nije moguće (zadatak).
- Poželjno je da koreliranost varijabli bude bar 0.3. U suprotnom, broj faktora neće biti puno manji od broja varijabli.
- KMO index > 0.8, Kaiser-Meyer-Olkin statistic.

NEJEDINSTVENOST. ROTACIJE.

Teorem

Pretp: $m > 1$ *i* F *rješenje faktorskog modela* (1). *Tada je* $T^\tau F$ *rješenje faktorskog modela za svaku ortogonalnu matricu* T.

Dokaz.

$$\begin{aligned} X - \mu &= \Lambda F + \epsilon \\ &= \Lambda(TT^\tau)F + \epsilon \\ &= \underbrace{(\Lambda T)}_{\Lambda^*}\underbrace{(T^\tau F)}_{F^*} + \epsilon \end{aligned}$$

i vrijedi

$$E(F^*) = E(T^\tau F) = T^\tau E(F) = 0$$
$$Var(F^*) = Var(T^\tau F) = T^\tau Var(F) T = T^\tau T = I$$

FAKTORSKA ANALIZA U R-U

Rješava se jednadžba (po Λ)

$$\Sigma = \Lambda\Lambda^{\tau} + \Psi.$$

– `factanal()` u R-ovom paketu `stat`[12].
– `fa()` u `psych` paketu. Ima 5 alternativa: fm="minres", fm="pa", fm="wls", fm="gls" and fm="ml" (fm=factor method)

```
library(GPArotation)
factanal(data, rotation="varimax")
```

Rotacija nalazi novu bazu (faktore) koji se mogu bolje interpretirati. varimax, quartimax i equimax — ortogonalne, promax i oblimin — neortogonalne transformacije (faktori su korelirani).

[12] pomoću *maksimalne vjerodostojnosti (normalnost varijabli)*

Literatura

- Mulaik, S. A. (2009). Foundations of factor analysis (2nd ed.). Statistics in the Social and Behavioral Sciences. Boca Raton: Chapman & Hall/CRC.
- https://cran.r-project.org/web/packages/psych/vignettes/overview.pdf
- http://www.di.fc.ul.pt/~jpn/r/factoranalysis/factoranalysis.html
- https://stat.ethz.ch/ maathuis/teaching/spring14/Rscript-factor-analysis.R