

Correspondence Analysis

Lavoslav Čaklović

Ožujak 2012, H. Biometrijsko društvo

Sadržaj

- 1 Simple CA. Tablica kontingencije
 - Što s tablicom?
 - Rezultat analize
 - Malo matematike
- 2 Kanonska CA. Jačina moždanog udara
 - Klasifikacija MU
 - Kovarijable
 - Analiza
 - Interpretacija
 - Sev-skala

Hair-Eyes

Colour Hair	Eye			
	Brown	Blue	Hazel	Green
Black	68	20	15	5
Brown	119	84	54	29
Red	26	17	14	14
Blond	7	94	10	16

Table: Tablica kontingencije: kosa-oči (boje).

U tablicu kontingencije¹ za svaki par vrijednosti kategorijskih varijabli (ξ_i, η_j) zapisujemo broj $f_{ij} \geq 0$ koji može biti indikator (binarna vrijednost) ili učestalost događaja koje obje kategorije uzrokuju.

¹Još jedan (engleski) naziv je **co-occurrence matrix**.

Cilj analize je...

...poredati klase u nekom logičkom slijedu pomoću pridruženih im težina: x_i — retka, y_j — stupca. (Jedna) logična veza među njima

$$x_i \propto \frac{\sum_j f_{ij} y_j}{\sum_j f_{ij}}, \quad y_j \propto \frac{\sum_i x_i f_{ij}}{\sum_i f_{ij}} \quad (1)$$

ili matrično zapisano

$$x \propto D^{-1} F y, \quad y \propto E^{-1} F^T x \quad (2)$$

gdje je F tablica kontingencije, F^T transponirana matrica, a

$$D = \text{diag}(f_{i\bullet}), \quad f_{i\bullet} = \sum_j f_{ij} \quad (\text{masa retka}),$$

$$E = \text{diag}(f_{\bullet j}), \quad f_{\bullet j} = \sum_i f_{ij} \quad (\text{masa stupca}).$$

Vektori x i y su rješenja jednadžbi

$$x \propto D^{-1}FE^{-1}F^T x, \quad y \propto E^{-1}F^T D^{-1}Fy, \quad (3)$$

Perronov korijen: $\lambda_{\max} = 1$, Perronov vektor: $\mathbf{1} = [1, 1, \dots, 1]^T$
(provjera). Tražimo netrivialna rješenja $\neq \mathbf{1}$.

Vektori x i y su rješenja jednadžbi

$$x \propto D^{-1}FE^{-1}F^T x, \quad y \propto E^{-1}F^T D^{-1}Fy, \quad (3)$$

Perronov korijen: $\lambda_{\max} = 1$, Perronov vektor: $\mathbf{1} = [1, 1, \dots, 1]^T$
(provjera). Tražimo netrivialna rješenja $\neq \mathbf{1}$.

Iterativni postupak (metoda potencije):

- 1 Starta se s proizvoljnim vektorom $y = y_0 \neq \mathbf{1}$.
- 2 Propisno ga standardiziramo² tako da vrijedi
 $\sum_j f_{\bullet j} y_j = 0, \quad \sum_j f_{\bullet j} y_j^2 = 1$.
- 3 Neka je $x = D^{-1}Fy$ i x standardiziramo:
 $\sum_i f_{i\bullet} x_i = 0, \quad \sum_i f_{i\bullet} x_i^2 = 1$.
- 4 Neka je $y = E^{-1}F^T x$.
- 5 Ponavljaju se postupci (2) – (4) do zadovoljavajuće točnosti.

Vektori x i y su rješenja jednadžbi

$$x \propto D^{-1}FE^{-1}F^T x, \quad y \propto E^{-1}F^T D^{-1}Fy, \quad (3)$$

Perronov korijen: $\lambda_{\max} = 1$, Perronov vektor: $\mathbf{1} = [1, 1, \dots, 1]^T$
(provjera). Tražimo netrivialna rješenja $\neq \mathbf{1}$.

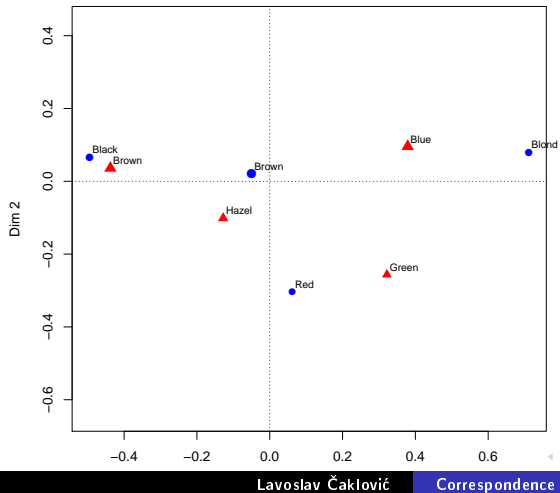
Iterativni postupak (metoda potencije):

- 1 Starta se s proizvoljnim vektorom $y = y_0 \neq \mathbf{1}$.
- 2 Propisno ga standardiziramo² tako da vrijedi
 $\sum_j f_{\bullet j} y_j = 0, \quad \sum_j f_{\bullet j} y_j^2 = 1$.
- 3 Neka je $x = D^{-1}Fy$ i x standardiziramo:
 $\sum_i f_{i\bullet} x_i = 0, \quad \sum_i f_{i\bullet} x_i^2 = 1$.
- 4 Neka je $y = E^{-1}F^T x$.
- 5 Ponavljaju se postupci (2) – (4) do zadovoljavajuće točnosti.

²Bez standardizacije postupak bi dao svojstveni vektor koji pripada maksimalnoj svojstvenoj vrijednosti 1, a to je $\mathbf{1}$.

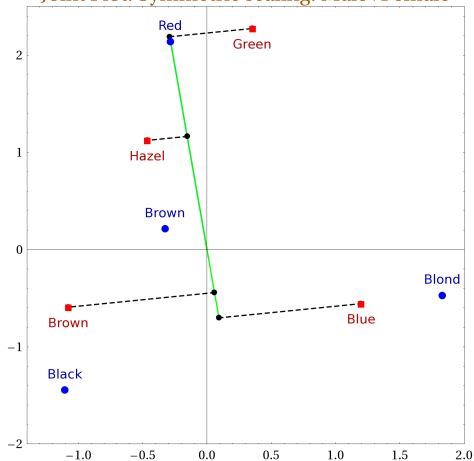
Biplot (R)

Hair Color and Eye Color



Biplot (Mathematica)

Joint Plot. Symmetric scaling. Male+Female



R kôd

```
library(ca)
library(vcd)
data(HairEyeColor)
HairEyeColor
(HairEye <- margin.table(HairEyeColor, c(1,2)))
# Male+Female

res<-ca(HairEye)
plot(res, map="symmetric",main="Hair Color and Eye Color
\nmap=symmetric", mass = TRUE)
title(xlab="Dim 1", ylab="Dim 2")
```

summary()

```
> summary(res)
```

```
Principal inertias (eigenvalues):
```

dim	value	%	cum%	scree plot
1	0.208773	89.4	89.4	*****
2	0.022227	9.5	98.9	**
3	0.002598	1.1	100.0	

```
-----  
Total: 0.233598 100.0
```

```
Rows:
```

	name	mass	qlt	inr	k=1	cor	ctr	k=2	cor	ctr
1	Blck	182	990	237	-505	838	222	215	152	379
2	Brwn	483	906	53	-148	864	51	-33	42	23
3	Red	120	945	65	-130	133	10	-320	812	551
4	Blnd	215	1000	646	835	993	717	70	7	47

Matematika CA

$$\text{Cov}(x, y) = \frac{1}{N} x^T F y \longrightarrow \max \quad (4)$$

$$u_n^T D x = 0, \quad u_m^T E y = 0 \quad (\text{centriranje}) \quad (5)$$

$$x^T D x = N, \quad y^T E y = N \quad (\text{normiranje}). \quad (6)$$

Matematika CA

$$\text{Cov}(x, y) = \frac{1}{N} x^T F y \longrightarrow \max \quad (4)$$

$$u_n^T D x = 0, \quad u_m^T E y = 0 \quad (\text{centriranje}) \quad (5)$$

$$x^T D x = N, \quad y^T E y = N \quad (\text{normiranje}). \quad (6)$$

Stacionarne jednadžbe su:

$$F y = \xi_x D x + \mu_x D u_n \quad (7)$$

$$F^T x = \xi_y E y + \mu_y E u_m, \quad (8)$$

gdje su $\xi_x, \xi_y, \mu_x, \mu_y$ Lagrangeovi multiplikatori.

Svođenje na SVD

Uvodimo supstitucije:

$$x =: D^{-\frac{1}{2}}p, \quad y =: E^{-\frac{1}{2}}q, \quad Z := D^{-\frac{1}{2}}FE^{-\frac{1}{2}}. \quad (9)$$

p i q rješavaju

$$Zq = \sigma p, \quad Z^T p = \sigma q, \quad (10)$$

što je ekvivalentno SVD dekompoziciji od Z

$$Z = P\Sigma Q^T. \quad (11)$$

Stupci od P , Q su oronormirani, oni su ujedno svojstveni vektori od $Z^T Z$ i ZZ^T respektivno tj.

$$ZZ^T P = P\Sigma^2, \quad Z^T ZQ = Q\Sigma^2. \quad (12)$$

Problem svojstvenih vrijednosti (3) ekvivalentan je problemu (12).

Jačina moždanog udara

Klasifikacija prema *jačini*:

- jak (Sev) — ako pacijent ima > 13 bodova na *NIHSS*³ skali.
- umjerena (Mod) — između 5 i 12 (uključivo i granice)
- blag (Mld) — ako pacijent ima < 5 bodova.

³The National Institutes of Health Stroke Scale

Jačina moždanog udara

Klasifikacija prema **jačini**:

- jak (Sev) — ako pacijent ima > 13 bodova na *NIHSS*³ skali.
- umjerena (Mod) — između 5 i 12 (uključivo i granice)
- blag (Mld) — ako pacijent ima < 5 bodova.

Klasifikacija prema **manifestaciji u početku**:

- glavoboljom (Hdc),
- epileptičkim napadajem (Mld),
- gubitkom svijesti (Con)
- nespecifičan (Unk)

³The National Institutes of Health Stroke Scale

Jačina moždanog udara

Klasifikacija prema **jačini**:

- jak (Sev) — ako pacijent ima > 13 bodova na *NIHSS*³ skali.
- umjerena (Mod) — između 5 i 12 (uključivo i granice)
- blag (Mld) — ako pacijent ima < 5 bodova.

Klasifikacija prema **manifestaciji u početku**:

- glavoboljom (Hdc),
- epileptičkim napadajem (Mld),
- gubitkom svijesti (Con)
- nespecifičan (Unk)

Zadatak: Opisati povezanost **Početak** → **JačinaMU**

³The National Institutes of Health Stroke Scale

Podaci s kovarijablama

	Jačina MU				Row covariates			
	Mld	Mod	Sev		Con	Hdc	Epi	Unk
s2	50	78	57	s2	0	0	0	1
s3	1	4	4	s3	0	0	1	0
s5	29	26	4	s5	0	1	0	0
s7	1	0	0	s7	0	1	1	0
s9	0	5	19	s9	1	0	0	0
s11	0	0	4	s11	1	0	1	0

Table: Klasifikacija moždanog udara moderirana njegovim početkom.

Za razliku od jednostavne CA, kategorije su kod kanonske CA određene ko-varijablama (dummy, indikatorske).

```
> res <- anacor(table, row.covariates = row.covariates,  
+ scaling = c("Benzecri","Benzecri"))
```

Chi-Square decomposition:

	Chisq	Proportion	Cumulative Proportion
Component 1	56.038	0.932	0.932
Component 2	3.090	0.051	0.984

```
> res$col.scores[,1] # Scaled col scores  
  Mld   Mod   Sev  
-0.6986 -0.2401 0.9514
```

```
> res$row.scores[,1] # Scaled row scores.  
  s2    s3    s5    s7    s9    s11  
0.0069 0.5350 -0.8794 -0.6035 1.6175 1.8934
```

Transformation plot (reci) i biplot

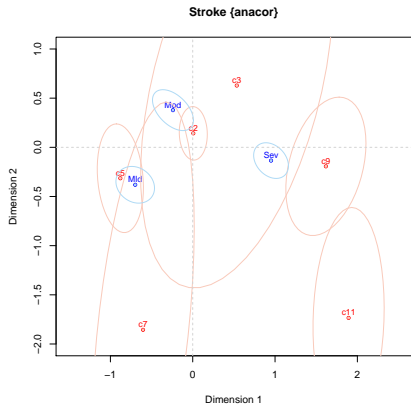
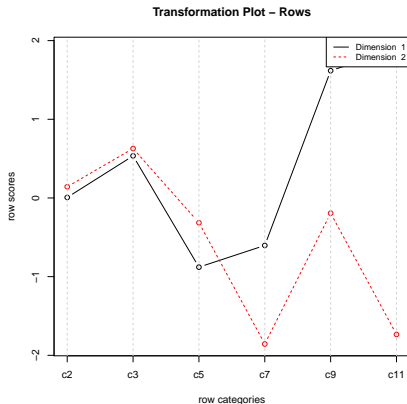
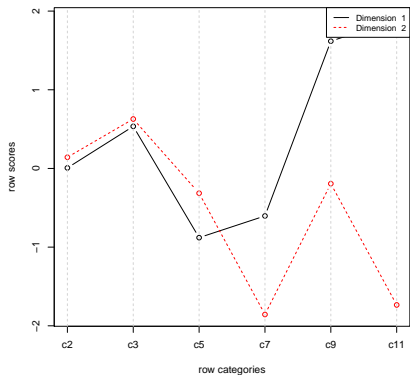


Figure: Transformation plot za tipove simptoma (retke) i jačine moždanog udara (stupce).

Transformation Plot – Rows



```
> res$left.singvec
```

```

                [,1]      [,2]
[1,]  0.07848346  0.2035717
[2,]  0.72025354 -0.6076077
[3,] -0.66417529 -0.7128864
[4,]  0.17452062 -0.2519706
[5,] -0.05906704  0.1329482
    
```

Con → [2,1], Hdc → [3,1]

Epi → [4,1], Unk → [5,1]

Figure: Transformation plot za tipove simptoma (retke) i jačine moždanog udara (stupce).

Sev-skala

Sev-skala je dobivena projiciranjem koordinata redaka (c2–c11) na Sev-pravac (koji prolazi ishodištem).

Sev-skala (res\$row.scores[,1])					
s11	s9	s3	s2	s7	s5
<i>Con + Epi</i>	<i>Con</i>	<i>Epi</i>	<i>Unk</i>	<i>Hdc + Epi</i>	<i>Hdc</i>
1.893	1.618	0.535	0.007	-0.604	-0.879
10	9	5	3	1	0
42	38	21	13	4	0

Table: Skala jačine moždanog udara ovisno o njegovom početku.