

Analýza kategorizovaných dat

Úkol číslo 3 – log-lineární analýza

Na datech z Pohybu obyvatelstva (1981 až 2002) o počtu rozvodů jsme provedli log-lineární analýzu s cílem nalézt nejvhodnější model pro odhalení struktury dat¹. V programu LEM jsme postupně vytvořili několik modelů od saturovaného, přes model podmíněné nezávislosti (R M Z) až po model plné dvojrozměrné interakce (testující výskyt vztahu mezi vzděláním muže a ženy v jednotlivých letech).

Podle statistických kritérií maximální věrohodnosti, BIC a delta se ukázal jako nejvhodnější a nejušpornější (d.f. = 39), model podmíněné nezávislosti (viz tabulka 1 – model č. 2). Dále statisticky vyhověly modely č. 4 a 5, které ovšem při vysvětlování vztahu mezi vzděláním muže a ženy nebyly nijak úsporné. Podle modelu 2 by tedy mezi vzděláním a rozvodovostí neměla existovat větší interakce.

Tabulka 1: Vybrané odhadované modely log-lineární analýzy

	Model	D.f.	L ²	delta	BIC
1	saturovaný	0	0	0	0
2	R M Z	39	1,99	7,6	-158,85
3	RM RZ	27	159,3	6,28	-83
4	RM RZ MZ	18	46,1	3,63	-110,9
5	RM RZ spe(MZ,1a,R,b)	16	33,43	3,13	-106,15

Abychom otestovali vliv časové dimenze, použili jsme k interpretaci model 5. Statistické výstupy modelu ukazují na to, že se charakteristika rozvodovosti ve všech třech sledovaných rocích mění (viz výsledky modelu). Z důvodu nízkého počtu let zahrnutých do analýzy nelze usuzovat na žádnou tendenci.

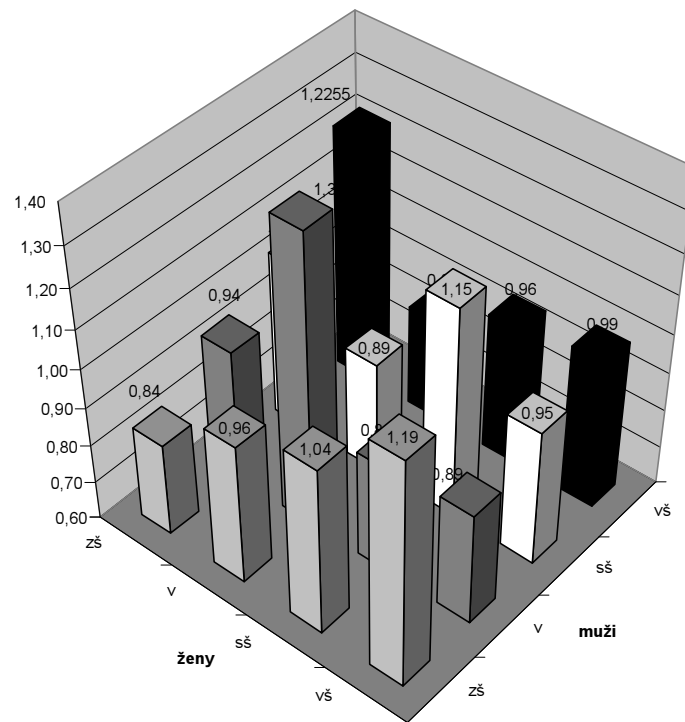
Výsledky modelu 5:

```
R [spe(MZ, 1a)]
1980      1.0000
1991      0.5813
2001      1.5777
```

¹ Počty rozvodů jsme vztáhli k počtu manželství v daném roce podle kombinace vzdělání partnerů a vynásobili konstantou 10 000.

Pro interpretaci samotné charakteristiky rozvodovosti jsme vybrali model 4, který též statisticky odpovídá datům (byť s d.f. 18). V tomto modelu je vliv času konstantní. Z grafu 1 je patrné, že s rostoucí hypergamií ženy provdané za muže se základním vzděláním riziko rozvodu výrazně stoupá. U mužů se tak děje analogicky. Dále je riziko rozvodu vyšší u homogamických manželství středoškoláků a vyučených.

Graf 1: Exponované koeficienty beta pro jednotlivá políčka tabulky



Použitá syntax

model 2

```
man 3
dim 3 4 4
lab R M Z *Rok Muž Žena
mod {R M Z}

dat [
111.85      153.82 250.89 315.96
81.51       121.59 112.46 125.26
81.71       71.15  128.33 132.82
64.61       56.66  101.85 119.03

107.77     190.66 263.05 206.93
87.30      140.49 127.45 130.68
81.19      89.83  130.12 114.22
77.14      73.45  95.01  95.87

132.82     174.81 204.49 344.29
68.99      177.36 126.58 132.60
61.10      87.67  156.39 113.28
83.06      73.65  93.33  99.45]
```

```

X-squared          = 200.8928 (0.0000)
L-squared          = 192.8133 (0.0000)
Cressie-Read      = 197.8991 (0.0000)
Dissimilarity index = 0.0760
Degrees of freedom = 39
Log-likelihood     = -23363.76744
Number of parameters = 8 (+1)
Sample size       = 6170.5
BIC(L-squared)    = -147.5608
AIC(L-squared)    = 114.8133
BIC(log-likelihood) = 46797.3552
AIC(log-likelihood) = 46743.5349

```

model 4

man 3

```

dim 3 4 4
lab R M Z *Rok Muž Žena
mod {RM RZ MZ}

```

```

X-squared          = 46.1218 (0.0003)
L-squared          = 46.2108 (0.0003)
Cressie-Read      = 46.1385 (0.0003)
Dissimilarity index = 0.0363
Degrees of freedom = 18
Log-likelihood     = -23290.46617
Number of parameters = 29 (+1)
Sample size       = 6170.5
BIC(L-squared)    = -110.8850
AIC(L-squared)    = 10.2108
BIC(log-likelihood) = 46834.0310
AIC(log-likelihood) = 46638.9323

```

model 5

man 3

```

dim 3 4 4
lab R M Z *Rok Muž Žena
mod {RM RZ spe(MZ, 1a, R, b)}

```

```

X-squared          = 33.4366 (0.0065)
L-squared          = 33.4833 (0.0064)
Cressie-Read      = 33.4448 (0.0064)
Dissimilarity index = 0.0313
Degrees of freedom = 16
Log-likelihood     = -23284.10243
Number of parameters = 31 (+1)
Sample size       = 6170.5
BIC(L-squared)    = -106.1574
AIC(L-squared)    = 1.4833
BIC(log-likelihood) = 46838.7586
AIC(log-likelihood) = 46630.2049

```