

Analýza kategorizovaných dat

Úkol číslo 3 – log-lineární analýza

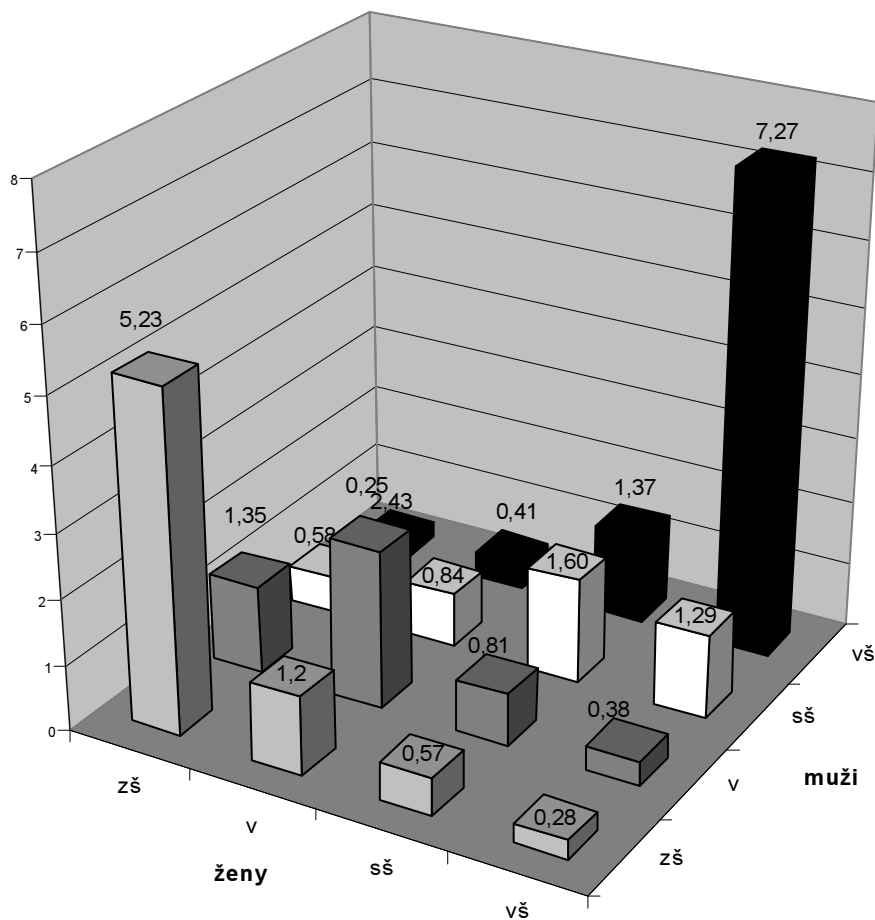
Na datech z Pohybu obyvatelstva (1981 až 2002) o počtu rozvodů jsme provedli log-lineární analýzu s cílem nalézt nejvhodnější model pro odhalení struktury dat. V programu LEM jsme postupně vytvořili několik modelů od saturovaného, přes model podmíněné nezávislosti (R M Z) až po model plné dvojrozměrné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

Podle statistických kritérií maximální věrohodnosti, BIC a delta se ukázal jako nejvhodnější, byť ne nejúspornější (d.f. = 16), model plné dvojrozměrné interakce (viz tabulka 1).

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	45403.9	28,3	44960.6
3	RM RZ	27	35680.1	26,6	35373.2
4	RM RZ MZ	18	256.0	1,9	51.4
5	{RM RZ MZ spe(MZ,1b,R,c,2)}	16	142.6	1,2	-39.2

Graf 1: Exponované koeficienty Beta pro model 5



Tyto koeficienty při hodnotě jedna odpovídají náhodnému rozložení. Při hodnotě vyšší než jedna (po odečtení jedničky) udávají kolikrát častěji se dané kategorie vyskytly oproti náhodnému rozložení. Analogicky pak při hodnotách nižších než jedna (kolikrát méně oproti náhodnému rozdělení).

Z grafu 1 je patrné, že se nejvíce rozvádějí homogamní manželství, z nich pak nejčastěji svazky vysokoškoláků a lidí se základním vzděláním. U párů vysoce heterogamních (pár muže VŠ, ženy ZŠ) je rozvodovost nejnižší – což pravděpodobně odráží skutečnost, že těchto svazků vzniká nejméně.

Testovali jsme také vliv časové dimenze a to v modelu 3. Statistické výstupy modelu ukazují na to, že charakteristika rozvodovosti se vlivem časové dimenze nemění.

Syntax

```
man 3
dim 3 4 4
lab R M Z *Rok Muž Žena
mod {RM RZ MZ spe(MZ, 1b, R, c, 2)}
des[0 1 2 ]
```

```
dat [
  6922 1243   785   60
  3913 5585  1984  127
  896  1056  2321  294
  147  233   942  708

  4055 1378   864   43
  3118 7464  3460  221
  783  1540  3419  410
  133  287   1243  948

  2026   1013   422   44
  1985   9979  3965  284
  383   1641  4597  531
  85     273   1287  1189]
```

*** STATISTICS ***

```
Number of iterations = 70
Converge criterion   = 0.0000008758

X-squared           = 148.5234 (0.0000)
L-squared           = 142.6557 (0.0000)
Cressie-Read       = 146.2377 (0.0000)
Dissimilarity index = 0.0120
Degrees of freedom  = 16
Log-likelihood      = -285004.86201
Number of parameters = 31 (+1)
Sample size         = 86286.0
BIC (L-squared)    = -39.1910
AIC (L-squared)    = 110.6557
BIC (log-likelihood) = 570362.0521
AIC (log-likelihood) = 570071.7240
```