

Kateřina Škařupová  
43275  
[skarupov@fss.muni.cz](mailto:skarupov@fss.muni.cz)

Přemysl Maršík  
13477  
[13477@mail.muni.cz](mailto:13477@mail.muni.cz)

## **Analýza kategorizovaných dat v sociologii**

### **Úkol č. 3**

#### **Zadání:**

Analyzujte trojrozměrná data o uskutečněných rozvodech podle vzdělání rozvádějícího se muže a ženy v letech 1980, 1990 a 2001 v České republice.

Data mají rozměr 3 x 4 x 4, tedy Roky x Vzdělání muže x Vzdělání ženy, přičemž Roky mají varianty 1980, 1990 a 2001, Vzdělání muže je základní, vyučen, středoškolské a vysokoškolské a Vzdělání ženy je základní, vyučena, středoškolské a vysokoškolské.

Zodpovězte následující otázky:

Jaká je struktura asociace mezi rozvádějícími se muži a ženami podle vzdělání?

Jaký je trend v asociaci mezi rozvádějícími se muži a ženami podle vzdělání v čase?

#### **Řešení:**

Pokusili jsme se aplikovat následující modely, které vždy obsahovaly model podmíněné nezávislosti mezi: Rok x Vzdělání muže a Rok x Vzdělání ženy, neboť tyto parametry jsou zřejmě podmíněné externími vlivy. Sledujeme tedy především typ asociace mezi: Vzdělání muže x Vzdělání ženy a časovou charakteristiku této asociace.

1) Model symetrické asociace s vyblokovými diagonálními prvky (individuální faktory), heterogenní časová závislost.

\*\*\* INPUT \*\*\*

```
man 3
dim 3 4 4
lab R M Z *Rok Muž Žena
mod {RM RZ fac(RMZ,12) spe(MZ,3a,R,b)}
```

```
des
[1 0 0 0
 0 2 0 0
 0 0 3 0
 0 0 0 4

 5 0 0 0
 0 6 0 0
 0 0 7 0
 0 0 0 8

 9 0 0 0
 0 10 0 0
 0 0 11 0
 0 0 0 12
]
```

```
dat[
* data jsme pro úsporu místa vynechali
]
```

\*\*\* STATISTICS \*\*\*

```
Number of iterations = 171
Converge criterion   = 0.0000009938

X-squared           = 70.1952 (0.0000)
L-squared           = 68.8870 (0.0000)
Cressie-Read        = 69.7173 (0.0000)
Dissimilarity index = 0.0055
Degrees of freedom  = 4
Log-likelihood       = -284967.97766
Number of parameters = 43 (+1)
Sample size          = 86286.0
BIC(L-squared)       = 23.4253
AIC(L-squared)       = 60.8870
BIC(log-likelihood)  = 570424.6685
AIC(log-likelihood)  = 570021.9553
```

Tento model ponechal pouze 4 stupně volnosti a hodnota BIC nám napovídá, že není vhodný. Koneckonců zdánlivá symetrie v datech je spíše demograficky podmíněná.

2) Modelujeme celou tabulku v jednotlivých letech, závislost na roku heterogenní.

\*\*\* INPUT \*\*\*

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

```
dat[
* data jsme pro úsporu místa vynechali
]
```

\*\*\* STATISTICS \*\*\*

```
Number of iterations = 75
Converge criterion   = 0.0000007866

X-squared           = 125.1880 (0.0000)
L-squared           = 125.1438 (0.0000)
Cressie-Read        = 125.0601 (0.0000)
Dissimilarity index = 0.0104
Degrees of freedom  = 16
Log-likelihood       = -284996.10605
Number of parameters = 31 (+1)
Sample size          = 86286.0
BIC(L-squared)       = -56.7030
AIC(L-squared)       = 93.1438
BIC(log-likelihood)  = 570344.5402
AIC(log-likelihood)  = 570054.2121
```

Tento model vyhovuje. Příliš mnoho informací nám nedává, snad jen, že struktura rozvodovosti v České republice podle vzdělání obou manželů se v jednotlivých letech podobá až na faktor, který ovšem není monotónní v čase. Uvedeme tyto parametry:

```
      R [spe(MZ,1a)]
1          1.0000
2          0.9539
3          1.1178
```

Pro důkladnější studium rozvodovosti by bylo třeba mít data z více let.

3) Pro kontrolu jsme odhadli model konstantní asociace, neboli oproti modelu 2 předpokládáme homogenní časovou závislost.

```
*** INPUT ***
```

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

```
dat[
* data jsme pro úsporu místa vynechali
]
```

```
*** STATISTICS ***
```

```
Number of iterations = 10
Converge criterion   = 0.0000004793

X-squared           = 250.6755 (0.0000)
L-squared           = 256.0076 (0.0000)
Cressie-Read        = 252.2861 (0.0000)
Dissimilarity index = 0.0188
Degrees of freedom  = 18
Log-likelihood       = -285061.53796
Number of parameters = 29 (+1)
Sample size          = 86286.0
BIC(L-squared)       = 51.4300
AIC(L-squared)       = 220.0076
BIC(log-likelihood)  = 570452.6732
AIC(log-likelihood)  = 570181.0759
```

Tento model má sice víc stupňů volnosti, ale realitě neodpovídá. Přijmeme tedy model 2.