

Kateřina Škařupová  
43275  
skarupov@fss.muni.cz

Přemysl Maršík  
13477  
13477@mail.muni.cz

### **Analýza kategorizovaných dat v sociologii Úkol č. 3**

Tabulky normalizované rozvodovosti jsme modelovali pomocí faktorů. Strukturu jednotlivých faktorů se nám však nepodařilo vhodně vysvětlit.

Použili jsme model:

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

```
man 3
dim 3 4 4
lab R M Z *Rok Muž Žena
mod { fac(MZ,3,R,c), fac(MZ,3,R,b), fac(MZ,1,R,b),
      fac(MZ,2,R,b), fac(MZ,3,R,b), fac(MZ,2,R,b) }

des[
0 0 0 1
0 2 0 0
0 0 3 0
0 0 0 0

1 2 3 0
0 0 0 0
0 0 0 0
0 0 0 0

0 0 0 0
0 0 1 1
0 0 0 0
0 0 0 0

0 0 0 0
0 0 0 0
0 0 0 1
0 0 2 2

0 0 0 0
0 0 0 0
0 1 0 0
2 3 0 0

0 0 0 0
1 0 0 0
2 0 0 0
0 0 0 0
]

dat[
*1988

1715 2359 3848 4846
1250 1865 1725 1921
1253 1091 1968 2037
```

991 869 1562 1825

\*1995

1653 2924 4034 3174  
1339 2155 1955 2004  
1245 1378 1996 1752  
1183 1127 1457 1470

\*2001

2037 2681 3136 5280  
1058 2720 1941 2034  
937 1345 2398 1737  
1274 1130 1431 1525 ]

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

Number of iterations = 508  
Converge criterion = 0.0000009850  
  
X-squared = 107.5754 (0.0000)  
L-squared = 108.7606 (0.0000)  
Cressie-Read = 107.9543 (0.0000)  
Dissimilarity index = 0.0099  
Degrees of freedom = 17  
Log-likelihood = -356897.12866  
Number of parameters = 30 (+1)  
Sample size = 94635.0  
BIC(L-squared) = -86.0217  
AIC(L-squared) = 74.7606  
BIC(log-likelihood) = 714137.9908  
AIC(log-likelihood) = 713854.2573

Několik polí v tabulkách modelujeme individuálně, zbytek je rozdělen do nezávislých faktorů, které se s časem vyvíjí multiplikativně. Model lze za cenu ztráty stupňů volnosti ještě zlepšit, když i prvky druhého z faktorů modelujeme individuálně.

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

man 3  
dim 3 4 4  
lab R M Z \*Rok Muž Žena  
mod {fac(MZ,3,R,c), fac(MZ,3,R,c), fac(MZ,1,R,b),  
fac(MZ,2,R,b), fac(MZ,3,R,b), fac(MZ,2,R,b)}

des[  
0 0 0 1  
0 2 0 0  
0 0 3 0  
0 0 0 0  
  
1 2 3 0  
0 0 0 0  
0 0 0 0  
0 0 0 0

```
0 0 0 0
0 0 1 1
0 0 0 0
0 0 0 0
```

```
0 0 0 0
0 0 0 0
0 0 0 1
0 0 2 2
```

```
0 0 0 0
0 0 0 0
0 1 0 0
2 3 0 0
```

```
0 0 0 0
1 0 0 0
2 0 0 0
0 0 0 0
]
```

```
dat[
```

```
  *1988
```

```
1715 2359 3848 4846
1250 1865 1725 1921
1253 1091 1968 2037
991  869  1562 1825
```

```
  *1995
```

```
1653 2924 4034 3174
1339 2155 1955 2004
1245 1378 1996 1752
1183 1127 1457 1470
```

```
  *2001
```

```
2037 2681 3136 5280
1058 2720 1941 2034
937  1345 2398 1737
1274 1130 1431 1525 ]
```

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

```
Number of iterations = 5000
Converge criterion   = 0.0000918623

X-squared           = 46.5417 (0.0000)
L-squared           = 46.7827 (0.0000)
Cressie-Read        = 46.6186 (0.0000)
Dissimilarity index = 0.0059
Degrees of freedom   = 13
Log-likelihood       = -356866.13969
Number of parameters = 34 (+1)
Sample size          = 94635.0
BIC(L-squared)       = -102.1685
AIC(L-squared)       = 20.7827
BIC(log-likelihood) = 714121.8440
AIC(log-likelihood) = 713800.2794
```

Pokusíme-li se ještě nějaké faktory zjednodušit, směřujeme ke vhodnému modelu: necháme třetí faktor v čase homogenní.

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

```
man 3
dim 3 4 4
lab R M Z *Rok Muž Žena
mod {fac(MZ,3,R,c), fac(MZ,3,R,c), fac(MZ,1,R,a),
     fac(MZ,2,R,b), fac(MZ,3,R,b), fac(MZ,2,R,b)}
```

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

```
1 2 3 0
0 0 0 0
0 0 0 0
0 0 0 0
```

```
0 0 0 0
0 0 1 1
0 0 0 0
0 0 0 0
```

```
0 0 0 0
0 0 0 0
0 0 0 1
0 0 2 2
```

```
0 0 0 0
0 0 0 0
0 1 0 0
2 3 0 0
```

```
0 0 0 0
1 0 0 0
2 0 0 0
0 0 0 0
]
```

```
dat[
*1988
```

```
1715 2359 3848 4846
1250 1865 1725 1921
1253 1091 1968 2037
991 869 1562 1825
```

```
*1995
```

```
1653 2924 4034 3174
1339 2155 1955 2004
1245 1378 1996 1752
1183 1127 1457 1470
```

```
*2001
```

2037 2681 3136 5280  
1058 2720 1941 2034  
937 1345 2398 1737  
1274 1130 1431 1525 ]

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

Number of iterations = 5000  
Converge criterion = 0.0000927155  
  
X-squared = 63.8791 (0.0000)  
L-squared = 64.7955 (0.0000)  
Cressie-Read = 64.1763 (0.0000)  
Dissimilarity index = 0.0064  
Degrees of freedom = 15  
Log-likelihood = -356875.14608  
Number of parameters = 32 (+1)  
Sample size = 94635.0  
BIC(L-squared) = -107.0713  
AIC(L-squared) = 34.7955  
BIC(log-likelihood) = 714116.9412  
AIC(log-likelihood) = 713814.2922

Při vší dobré vůli se nám nepodařilo vytvořit lepší model.