PREHLIADKA PRÁC MLADÝCH ŠTATITIKOV A DEMOGRAFOV

MODELOVANIE ÚMRTNOSTNÝCH TABULIEK POMOCOU ZOVŠEOBECNENÉHO GOMPERTZOVHO-MAKEHAMOVHO MODELU

Nikolas Márkus

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

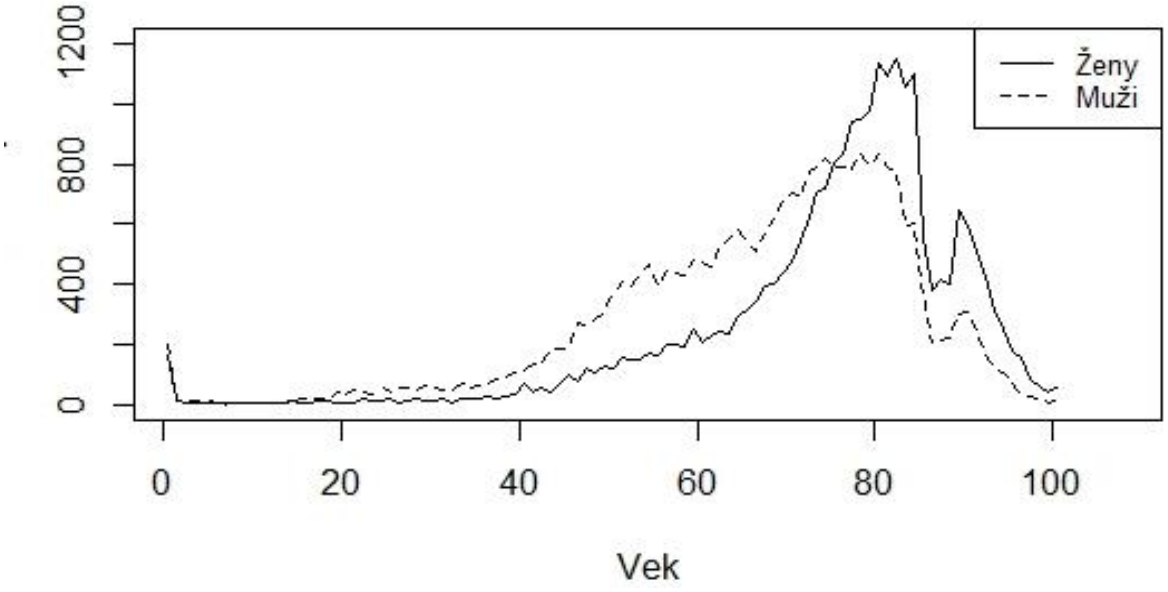
1. December 2016, Bratislava

Obsah

- základné charakteristické veličiny úmrtnostných tabuliek
- Gompertzov-Makehamov model
- proces fitovania úmrtnostných dát
- fitovanie na celom resp. skrátenom vekovom intervale
- výsledky a frekvencie najvhodnejších modelov

Charakteristiky úmrtnostných tabuliek

- x vyjadruje vek osoby, kde $x \in [0, \omega)$, pričom ω je predpokladaný najvyšší vek, ktorý môže dosiahnuť osoba sledovaného súboru (obvykle 100 rokov),
- P_x vyjadruje reálny počet žijúcich x-ročných osôb v strede pozorovaného obdobia,
- D_x vyjadruje reálny počet ľudí, ktorí zomreli ako x-roční, t. j. nedožili sa veku $x + 1$ rokov,



Charakteristiky úmrtnostných tabuliek

- m_x je charakteristická veličina, ktorá vyjadruje relatívny počet úmrtí na veľkosť sledovanej populácie vo veku x; nazývame ju obecnou mierou úmrtnosti a definujeme ju:

$$m_x = \frac{D_x}{P_x}.$$

- μ_x je intenzita úmrtnosti, ktorá predstavuje pravdepodobnosť, že osoba vo veku *x* zomrie v ľubovoľne krátkom časovom okamihu. Matematicky ju definujeme:

$$\mu_x = \lim_{dx \to 0} \frac{P(T_0 \leq x + dx \mid T_0 > x)}{dx},$$

kde T_x je budúca dĺžka osoby vo veku x.

Gompertzov-Makehamov (GM) model

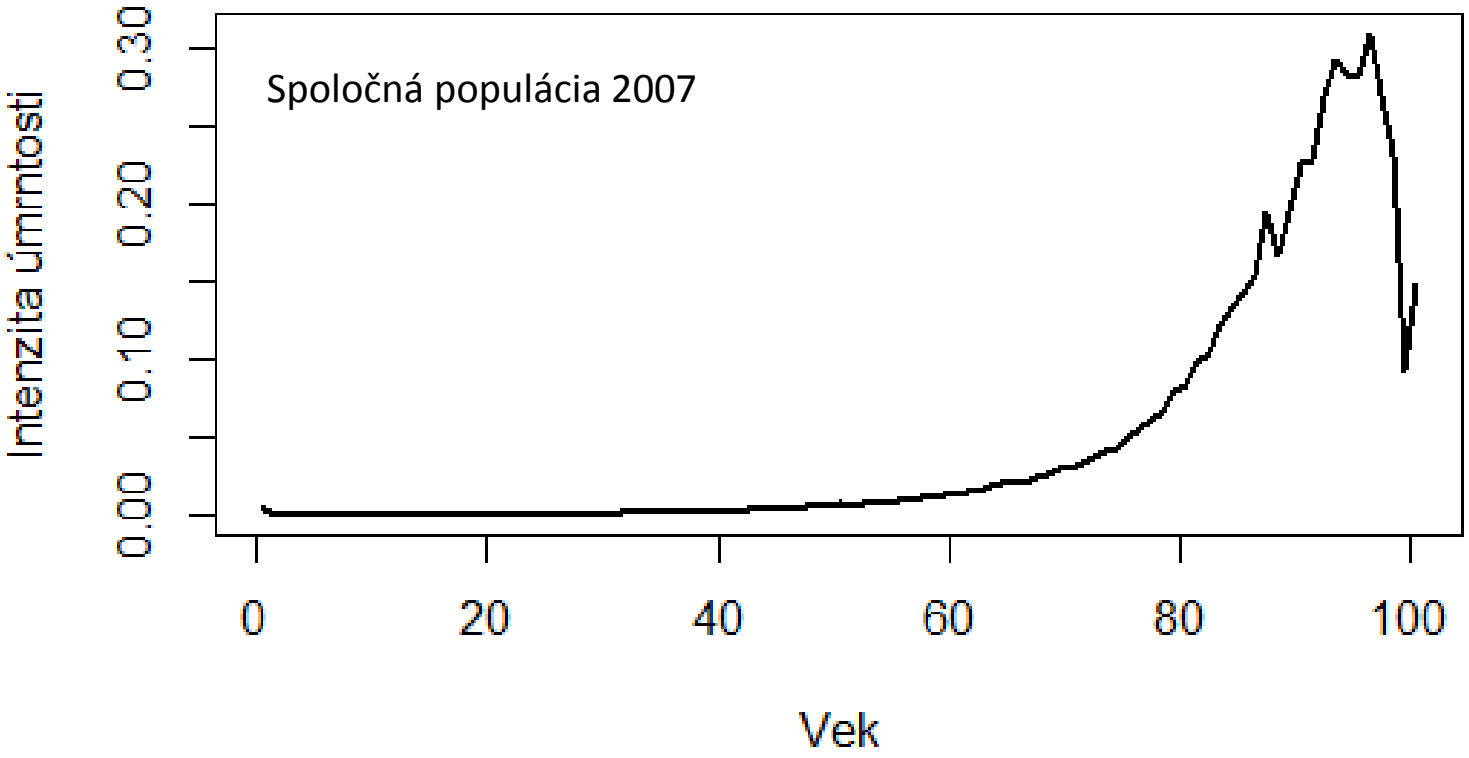
Základný GM model:

$$\mu_x = A + BC^x$$

Zovšeobecnený GM model:

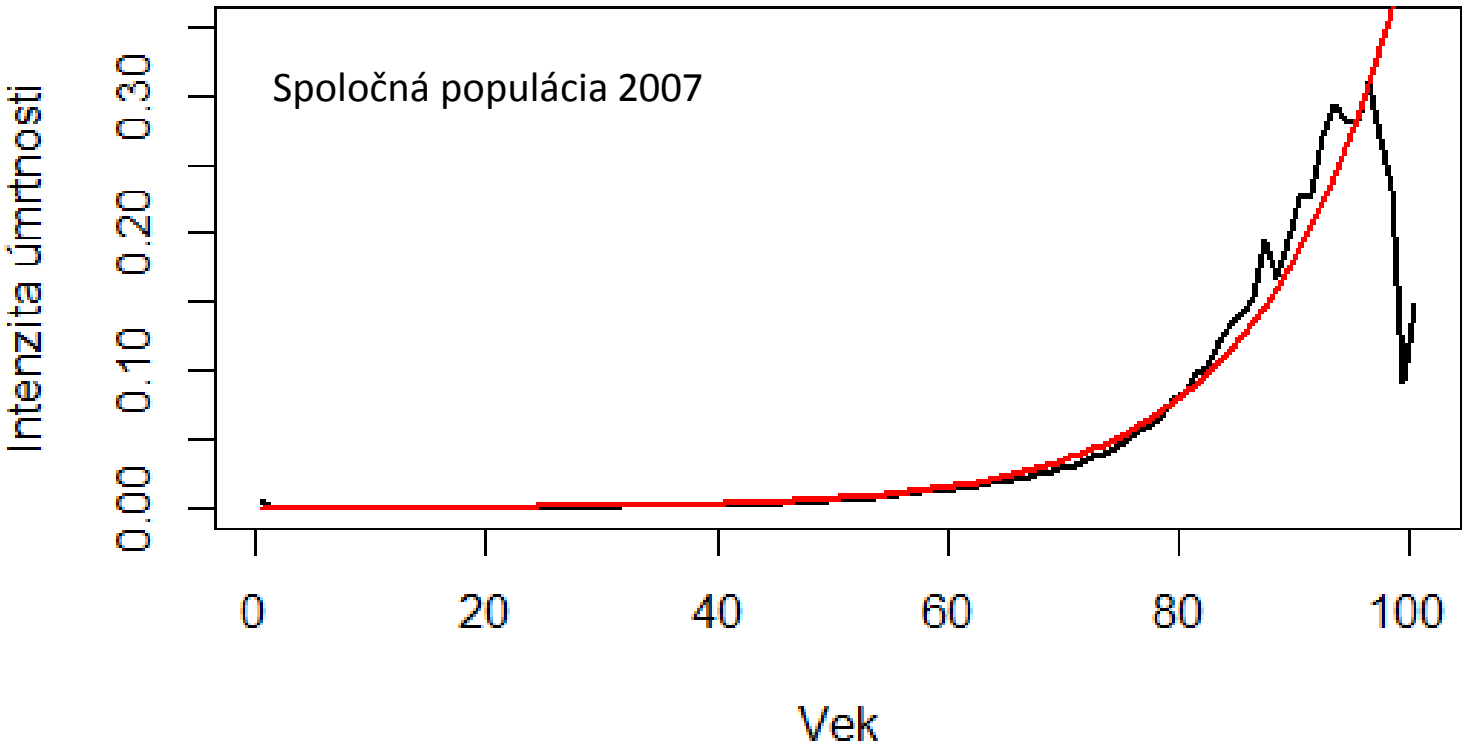
$$\mu_x = GM(r,s) = \sum_{i=0}^{r-1} \beta_{(i)} x^i + \exp\left(\sum_{j=0}^{s-1} \beta_{(r+j)} x^j\right)$$

Ukážka úmrtnostných dát/motivácia



Zdroj: Úmrtnostné tabuľky, Štatistický úrad Slovenskej republiky.

Ukážka úmrtnostných dát/motivácia



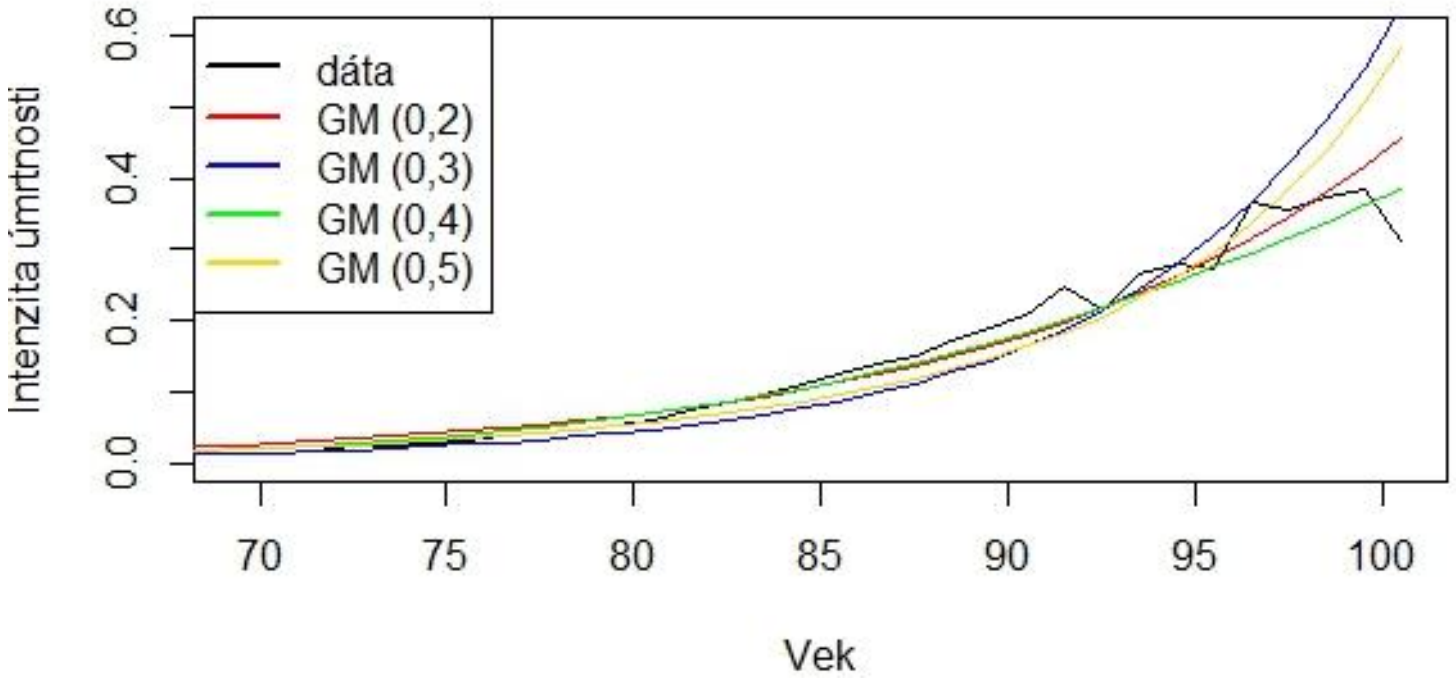
Zdroj: Úmrtnostné tabuľky, Štatistický úrad Slovenskej republiky.

Fitovanie na celom vekovom intervale

- Vekový interval : $x \in [0,100]$
- Rýdzo exponenciálne modely: $GM(0,s): \quad \mu_x = \mathrm{e}^{\beta_0+\beta_1 x+\beta_2 x^2+\cdots+\beta_{s-1}x^{s-1}}$
- Počiatočné odhady do optimalizácie boli vypočítané cez lineárnu regresiu.
  za pomoci logaritmickej transformácie: $\ln(\mu_x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_{s-1}x^{s-1}.$
- Formulácia úlohy: $$\min_{\beta \in \mathbb{R}^s} \sum_{x=0}^{100} |m_x - \mu_x|.$$
- Testovanie kvality modelu: vizuálny fit, počet parametrov.

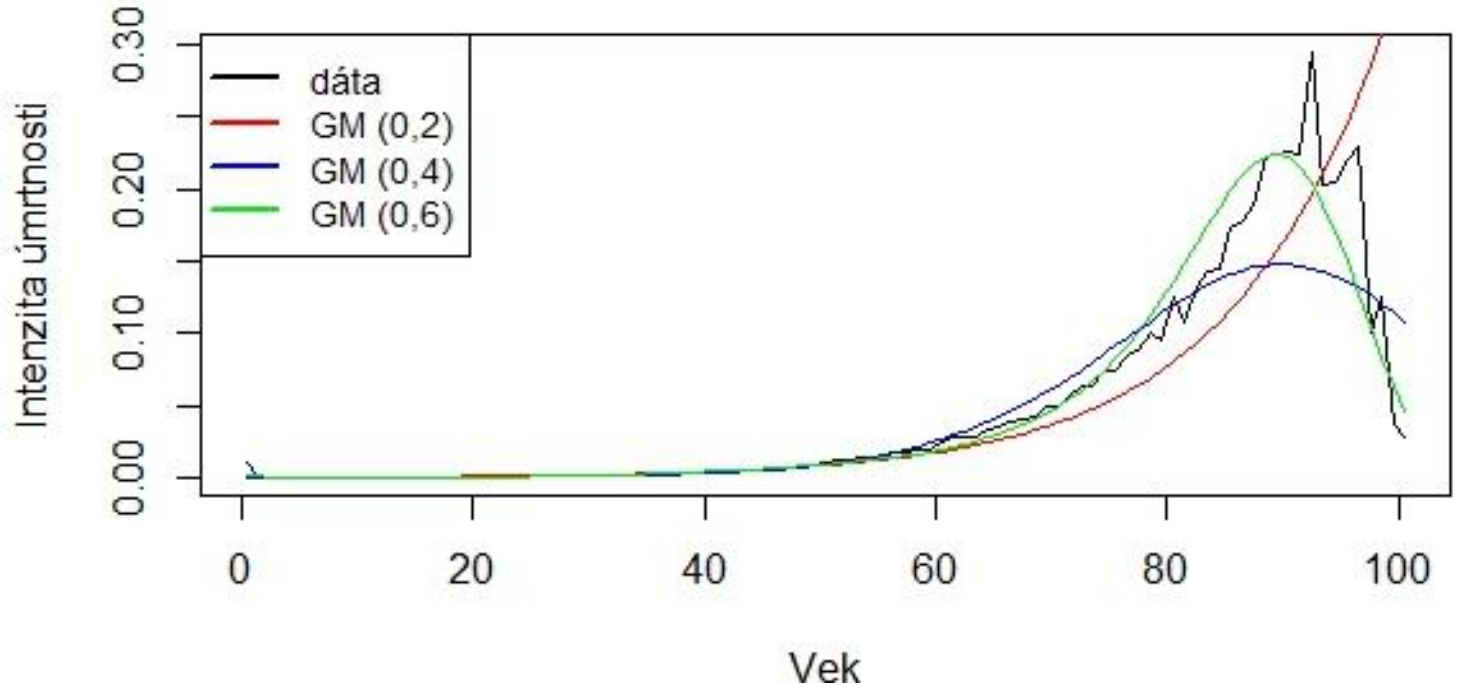
Fitovanie na celom vekovom intervale

Porovnanie vyšších a nižších rýdzo exponenciálnych GM modelov pri relatívne nízkych odchýlkach (ženská populácia, 2011) pre vekový interval 70 až 100 rokov.



Fitovanie na celom vekovom intervale

Porovnanie vyšších a nižších rýdzo exponenciálnych GM modelov pri extrémne vysokých odchýlkach (mužská populácia, 2000).



Výsledky fitovania na celom vekovom intervale

Rok	Ženy	Muži	Spoločné
1996	GM(0,5)	GM(0,3)	
1997	GM(0,2)	GM(0,2)	
1998	GM(0,5)	GM(0,3)	
1999	GM(0,2)	GM(0,2)	
2000	GM(0,2)	GM(0,3)	
2001	GM(0,4)	GM(0,2)	
2002	GM(0,5)	GM(0,2)	
2003	GM(0,4)	GM(0,2)	GM(0,2)
2004	GM(0,5)	GM(0,3)	GM(0,2)

Rok	Ženy	Muži	Spoločné
2005	GM(0,5)	GM(0,2)	GM(0,5)
2006	GM(0,2)	GM(0,2)	GM(0,2)
2007	GM(0,2)	GM(0,2)	GM(0,5)
2008	GM(0,2)	GM(0,2)	GM(0,2)
2009	GM(0,2)	GM(0,2)	GM(0,2)
2010	GM(0,2)	GM(0,2)	GM(0,3)
2011	GM(0,4)	GM(0,2)	GM(0,4)
2012	GM(0,4)	GM(0,2)	GM(0,4)
2013	GM(0,4)	GM(0,2)	GM(0,5)

Frekvencia najvhodnejších modelov:

Model	Ženy	Muži	Spoločné
GM(0,2)	8	14	5
GM(0,3)	0	4	1
GM(0,4)	5	0	2
GM(0,5)	5	0	3

Fitovanie na skrátenom vekovom intervale

- Vekový interval : $x \in [15,62]$

- Zmiešané modely GM(r, s)

- Počiatočná odhadnutá hodnota pre absolútny člen: $\beta_0 = min\{m_1, m_2, \ldots m_9, m_{10}\}$

- Formulácia úlohy:

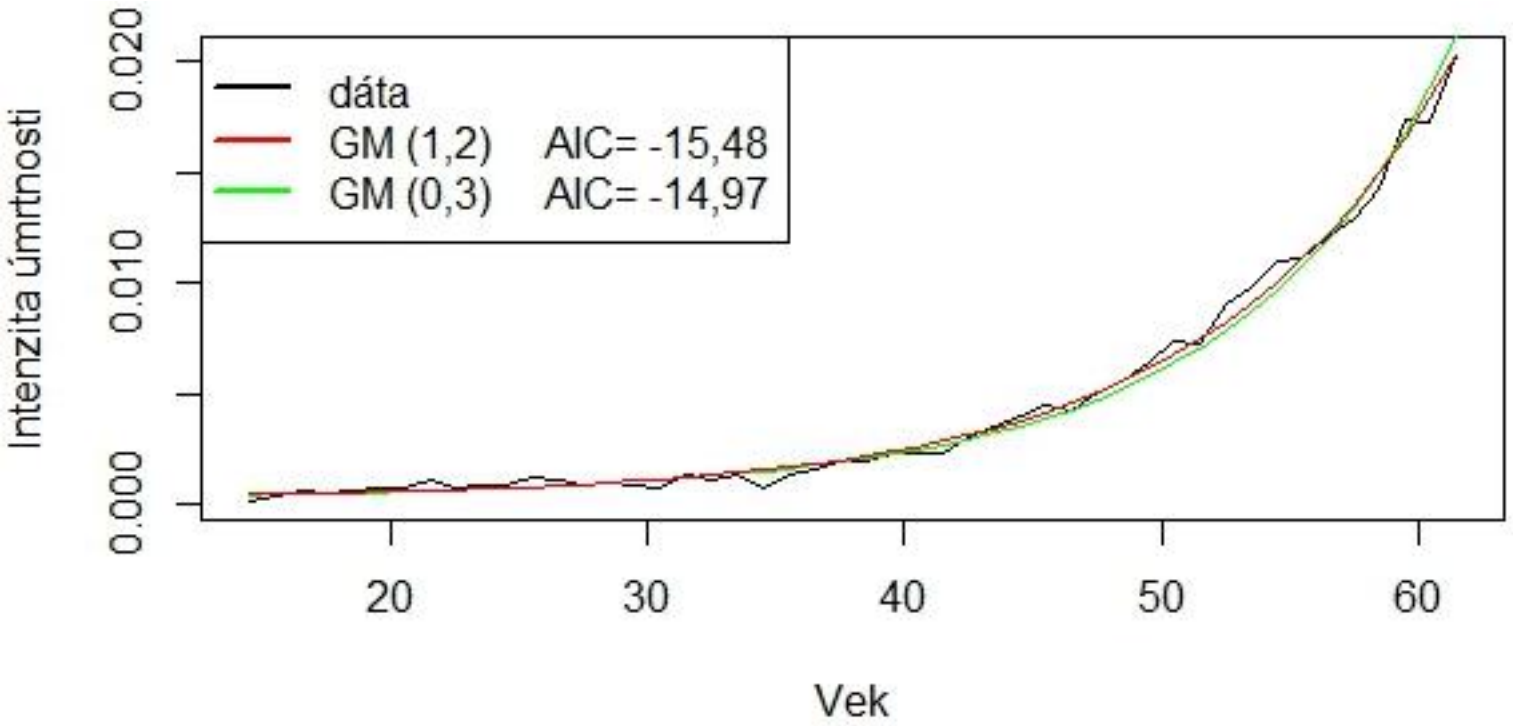
$$\min_{\beta \in \mathbb{R}^{r+s}} \sum_{x=15}^{62} |m_x - \mu_x|,$$

$$GM(r,s) \geq 0.$$

- Testovanie kvality modelu: Akaikeho informačné kritérium

Fitovanie na skrátenom vekovom intervale

Porovnanie najlepšieho exponenciálneho a zmiešaného GM modelu pri dátach pre mužskú populáciu z roku 2012.



Výsledky fitovania na skrátenom vekovom intervale

Rok	Ženy	Muži	Spoločné
1996	GM(0,2)	GM(0,2)	
1997	GM(0,4)	GM(1,2)	
1998	GM(0,4)	GM(0,2)	
1999	GM(0,4)	GM(1,2)	
2000	GM(0,4)	GM(1,2)	
2001	GM(0,4)	GM(1,2)	
2002	GM(0,4)	GM(0,2)	
2003	GM(0,4)	GM(1,2)	GM(0,2)
2004	GM(0,4)	GM(1,2)	GM(1,2)

Rok	Ženy	Muži	Spoločné
2005	GM(0,2)	GM(1,2)	GM(0,2)
2006	GM(0,4)	GM(0,2)	GM(0,2)
2007	GM(0,4)	GM(0,2)	GM(0,2)
2008	GM(0,4)	GM(0,2)	GM(0,2)
2009	GM(0,2)	GM(1,2)	GM(0,2)
2010	GM(0,2)	GM(0,2)	GM(0,2)
2011	GM(0,2)	GM(0,2)	GM(0,2)
2012	GM(0,4)	GM(1,2)	GM(0,2)
2013	GM(0,2)	GM(1,3)	GM(0,2)

Frekvencia najvhodnejších modelov:

Model	Ženy	Muži	Spoločné
GM(0,2)	6	8	10
GM(0,3)	0	0	0
GM(0,4)	12	0	0
GM(0,5)	0	0	0
GM(1,2)	0	9	1
GM(1,3)	0	1	0
GM(2,2)	0	0	0
GM(2,3)	0	0	0

Zhrnutie výsledkov

- na rôzne typy populácii sú vhodné rôzne modely,
- pre ženskú populáciu sú najlepšie vyššie exponenciálne modely,
- pre mužskú populáciu sú naopak nižšie modely vhodné (základný GM model),
- pre spoločnú populáciu postačuje GM(0,2) model,
- veľmi vysoké GM modely sú demograficky nepoužiteľné.

Ďakujem za pozornosť