

# METODIKA VÝPOČTU ÚMRTNOSTNÍCH TABULEK

## Základní metodické poznámky

Podrobné úmrtnostní tabulky jsou založeny na III. hlavních souborech demografických událostí. Vstupní pravděpodobnosti úmrtí jsou vypočteny nepřímou metodou, tj. odvozeny z pozorovaných specifických měr úmrtnosti (viz níže). Tabulky jsou podrobné, tj. s jednoletým věkovým intervalem, s nejvyšším věkovým intervalem 105 a více let. Jsou počítány odděleně pro muže a ženy.

S ohledem na vyloučení nahodilých výkyvů jsou úmrtnostní tabulky pro kraje (NUTS 3) zpracovány za dvouleté kalendářní období a úmrtnostní tabulky za okresy a správní obvody obcí s rozšířenou působností (SO ORP) za pětileté kalendářní období. Přesto nízké počty zemřelých a použitá (časově i prostorově jednotná) metodika vyrovnávání a modelování měr úmrtnosti mohou stát za výkyvy v časové řadě.

Základními vstupními daty jsou údaje o počtu zemřelých a středním stavu obyvatel daného věku a pohlaví v daném roce, kterými ČSÚ disponuje a používá ve svých výstupech. Do výpočtu přitom vstupují vstupní data pouze do jednotky věku, pro kterou je dostupný údaj jak o počtu zemřelých, tak o středním stavu obyvatel (například pro období let 1920–1924 byly jako vstupy využity údaje pouze pro věky 0–94 let včetně). Vstupní data vždy odpovídají územnímu členění platnému k 1. 1. daného roku události.

Český statistický úřad využívá zde popsanou metodiku zpracování úmrtnostních tabulek od roku 2018 (počínaje zpracováním údajů za rok 2017). Pro sledování dynamiky vývoje úmrtnosti však byl na podzim roku 2018 proveden shodnou metodikou také retrospektivní výpočet úmrtnostních tabulek v časové řadě, a to od roku 1920 v případě úmrtnostních tabulek pro celou Českou republiku, resp. od roku 2001 v případě úmrtnostních tabulek za regiony soudržnosti, kraje a okresy. Tyto retrospektivní tabulky byly vydány v publikaci Úmrtnostní tabulky v časové řadě (ČSÚ, 2018).

## Ukazatele úmrtnostních tabulek

**Počet zemřelých ( $D_x$ )** uvádí absolutní počet zemřelých ve věku  $x$  během daného období.

**Počet obyvatel ( $P_x$ )** uvádí střední stav obyvatel ve věku  $x$ , resp. v případě víceletých období uvádí součet středních stavů v příslušných letech. V případě tabulek za ČR, regiony soudržnosti, kraje a okresy je střední stav definován jako bilancovaný počet obyvatel k 1. 7. daného roku. V případě tabulek za SO ORP je střední stav vypočten jako průměr počátečního a koncového stavu daného roku.

**Pravděpodobnost úmrtí ( $q_x$ )** vyjadřuje pravděpodobnost, že osoba dožívající se přesného věku  $x$  let v daném období (tj. před dosažením věku  $x+1$ ) zemře:

$$q_x = \frac{m_x}{1 + (1 - a_x) \cdot m_x}$$

kde  $m_x$  označuje míru úmrtnosti v daném věku a parametr  $a_x$  vyčísluje průměrný počet člověkoroků prožitých v daném věkovém intervalu zemřelými jedinci (viz níže).

**Tabulkový počet dožívajících ( $l_x$ )** je hypotetický počet osob, které se dožijí věku  $x$  let z 100 000 živě narozených (kořen tabulky  $l_0 = 100\,000$ ) při zachování řádu úmrtnosti sledovaného období:

$$l_{x+1} = l_x \cdot (1 - q_x)$$

**Tabulkový počet zemřelých ( $d_x$ )** vyjadřuje hypotetický počet zemřelých osob v dokončeném věku  $x$  let; je počítán jako rozdíl dvou po sobě jdoucích tabulkových počtů dožívajících:

$$d_x = l_x - l_{x+1}$$

**Tabulkový počet žijících ( $L_x$ )** je hypotetický počet člověkoroků prožitých mezi přesnými věky  $x$  a  $x+1$ :

$$L_x = l_x - (1 - a_x) \cdot d_x$$

**Pomocný ukazatel ( $T_x$ )** vyjadřuje počet let života, které má tabulková populace (nikoliv jednotlivec) v daném věku  $x$  ještě před sebou, a je dán kumulací počtu žijících  $L_x$  od věku  $x$  do nejvyššího věku tabulky:

$$T_x = \sum_x^{105+} L_x$$

**Střední délka života** neboli **naděje dožití ( $e_x$ )** udává průměrný počet let, který prožije právě  $x$ -letá osoba při zachování úmrtnostních poměrů daného období. Jedná se o syntetický ukazatel, který odráží úmrtnostní poměry ve všech věkových skupinách.

$$e_x = \frac{T_x}{l_x}$$

## Podrobný postup konstrukce úmrtnostní tabulky

### A. Zpracování vstupních dat

#### 1. Výpočet pozorovaných měr úmrtnosti

Z empirických dat je vypočtena tzv. reálná (pozorovaná) míra úmrtnosti ( $m_x$ ), jakožto podíl zemřelých a středního stavu obyvatel v daném věku, pohlaví a územní jednotce:

$${}_t m_x^{\text{ČR, region soudržnosti}} = \frac{{}_t D_x}{1.7.(t) P_x}$$

$${}_{t-1,t} m_x^{\text{kraj}} = \frac{{}_{t-1} D_x + {}_t D_x}{1.7.(t-1) P_x + 1.7.(t) P_x}$$

$${}_t m_x^{\text{okres}} = \frac{\sum_t {}_t D_x}{\sum_t 1.7.(t) P_x}$$

$${}_t m_x^{\text{SO ORP}} = \frac{\sum_t {}_t D_x}{\sum_t \left( \frac{1.1.(t) P_x + 31.12.(t) P_x}{2} \right)}$$

#### 2. Vyrovnání pozorovaných měr úmrtnosti

Pro ošetření možných náhodných fluktuací reálných (pozorovaných) měr úmrtnosti, které jsou obzvláště běžné ve věcích s nízkou úmrtností nebo s nízkými počty žijících, jsou míry úmrtnosti vyhlazeny metodou zobecněných adaptivních modelů v kombinaci s tzv. P-spliny, někdy označováno jako metoda P-GAM (Eilers a Marx, 1996). Prediktorem je věk, predikovanou hodnotou počet zemřelých, počet obyvatel je použit jako offset. U počtů zemřelých se předpokládá Poissonovo rozdělení. Míra vyhlazení je „adaptována“ na počty zemřelých v daném věku, tzn. nižší počty zemřelých jsou vyhlazovány více než vysoké počty zemřelých, kde je tak zachována významnost původního jevu/trendu.

#### 3. Modelování úmrtnosti ve vysokém věku

S ohledem na nízké počty zemřelých ve vysokém věku a nižší spolehlivost údajů o středním stavu nejstarší populace jsou vyrovnané pozorované míry úmrtnosti v nejvyšších věcích nahrazeny modelovými hodnotami. Úmrtnostní modely přitom obecně pracují se spojitou definicí míry úmrtnosti, tzv. force of mortality (někdy překládáno jako síla úmrtnosti), značenou  $\mu_x$ . V praxi se předpokládá, že při monotónním nárůstu funkce úmrtnosti s věkem přibližně platí následující vztah mezi silou a mírou úmrtnosti v daném věku (Thatcher a kol., 1998):

$$m_x \cong \mu_{x+0,5}$$

Zde zvolený model je založen na logistické křivce, která zohledňuje zpomalení nárůstu úmrtnosti s věkem. Konkrétně jde o model, který poprvé použil Kannisto (Thatcher a kol., 1998). Výpočtový program tak modeluje věkově specifickou úmrtnost pomocí funkce:

$$\mu_x = \frac{\alpha e^{\beta x}}{1 + \alpha e^{\beta x}} + c$$

kde  $x$  je věk a  $\alpha$ ,  $\beta$  jsou jednotlivé parametry,  $c$  je konstanta. Parametry funkce jsou odhadnuty metodou maximální věrohodnosti. Výchozí odhad parametrů je založen na předpokladu Poissonova rozdělení počtů zemřelých ( $D$ ) ve věku  $x$ . Na základě těchto předpokladů je pak maximalizován logaritmus věrohodnostní funkce ( $L$ ) ve tvaru:

$$\ln L = \sum_x D_x \ln(\mu_x) - E_x \mu_x$$

Při optimalizaci (záporného logaritmu) věrohodnostní funkce program pracuje s vlastními počátečními hodnotami parametrů. Pro odhad parametrů je nastaveno věkové rozpětí 70–90 let.

Při přechodu z vyrovnaných na modelové hodnoty se předpokládá, že věk, od kterého modelové hodnoty charakterizují úmrtnost lépe než vyrovnané hodnoty, je vyšší než 75 let. Pro hladký průběh nahrazení vyrovnaných hodnot modelovými hodnotami se nejprve hledá věk  $y$  vyšší než 75 let, kde je pozorován nejmenší rozdíl mezi vyrovnanou a modelovou mírou úmrtnosti, tzn. takový věk  $y$ , pro který platí:

$$|m_x^{vyrovn} - m_x^{model}| = \min, \quad y \geq 75$$

Po nalezení věku  $y$  je provedena postupná graduace vyrovnaných měř na modelové hodnoty pro věky  $x \in [(y - 4), (y + 4)]$  dle vztahu:

$$m_x^{grad} = \left(1 - \frac{x - y + 5}{10}\right) \cdot m_x^{vyrovn} + \left(\frac{x - y + 5}{10}\right) \cdot m_x^{model}$$

4. Pro výpočet úmrtnostní tabulky je použita následující kombinace hodnot věkově specifických měř úmrtnosti:

pro věk 0 let:	reálné (pozorované) hodnoty,
pro věk 1 až $(y - 5)$ let:	vyrovnané hodnoty,
pro věk $(y - 4)$ až $(y + 4)$ let:	graduace mezi vyrovnanými a modelovými hodnotami,
pro věk $(y + 5)$ let a vyšší:	modelové hodnoty.

## B. Převod reálných vyrovnaných dat na tabulkové

Základem pro výpočet úmrtnostní tabulky nepřímou metodou je převod věkově specifických měř úmrtnosti ( $m_x$ ) na pravděpodobnosti úmrtí ( $q_x$ ). Za předpokladu rovnosti reálné a tabulkové míry úmrtnosti a za předpokladu stability měř úmrtnosti v hypotetické tabulkové kohortě používáme následující převodní rovnici (Chiang, 1984):

$$q_x = \frac{m_x}{1 + (1 - a_x) \cdot m_x}$$

Parametr  $a_x$  vyčísluje průměrný počet člověkoroků prožitých v daném věkovém intervalu zemřelými jedinci. S ohledem na obecně platný předpoklad rovnoměrného rozložení přesného věku zemřelých je použita hodnota 0,5 pro všechny věky s výjimkou věku 0 a posledního (otevřeného) věkového intervalu.

Pro odhad průměrného věku zemřelých kojenců ( $a_0$ ) je (počínaje daty za rok 1986) využít nejpresnější způsob výpočtu, kdy je hodnota  $a_0$  získána přímo z reálných individuálních dat jakožto pozorovaný podíl průměrného věku zemřelých kojenců (ve dnech) na celkovém trvání kalendářního roku (Andreev a Kingkade, 2015):

$$a_0^{poz} = \left( \frac{\sum X_{D_{i,0}}^{(dny)}}{\sum D_{i,0}} \right) / 365$$

kde  $D_{i,0}$  je individuální úmrtí v kojeneckém věku a  $X_{D_{i,0}}^{(dny)}$  individuální věk zemřelého kojence ve dnech. Při zpracování úmrtnostních tabulek za ČR pro roky 1920 až 1985 byl ve vazbě na nedostatečný detail dostupných dat o zemřelých kojencích parametr  $a_0$  nastaven defaultně v hodnotě 0,1.

V případě nižších územních celků se vždy uvádí hodnota  $a_0$  vypočtená z dat za celou ČR v daném období.

U poslední otevřené věkové skupiny (zde 105+) je parametr  $a_x$  vypočten jako (viz např. Preston a kol., 2000):

$$a_{105+} = \frac{1}{m_{105+}}$$

Další postup výpočtu úmrtnostních tabulek navazuje na vypočtené hodnoty  $q_x$  (viz Ukazatele úmrtnostních tabulek).