

Teoretické a metodologické prístupy konštrukcie generačných úmrtnostných tabuliek a ich zhodnotenie v kontexte dátových podmienok Slovenska

Autor: Ján Mészáros

Teoreticko-metodologický materiál

Číslo čiastkovej úlohy 900-5/1

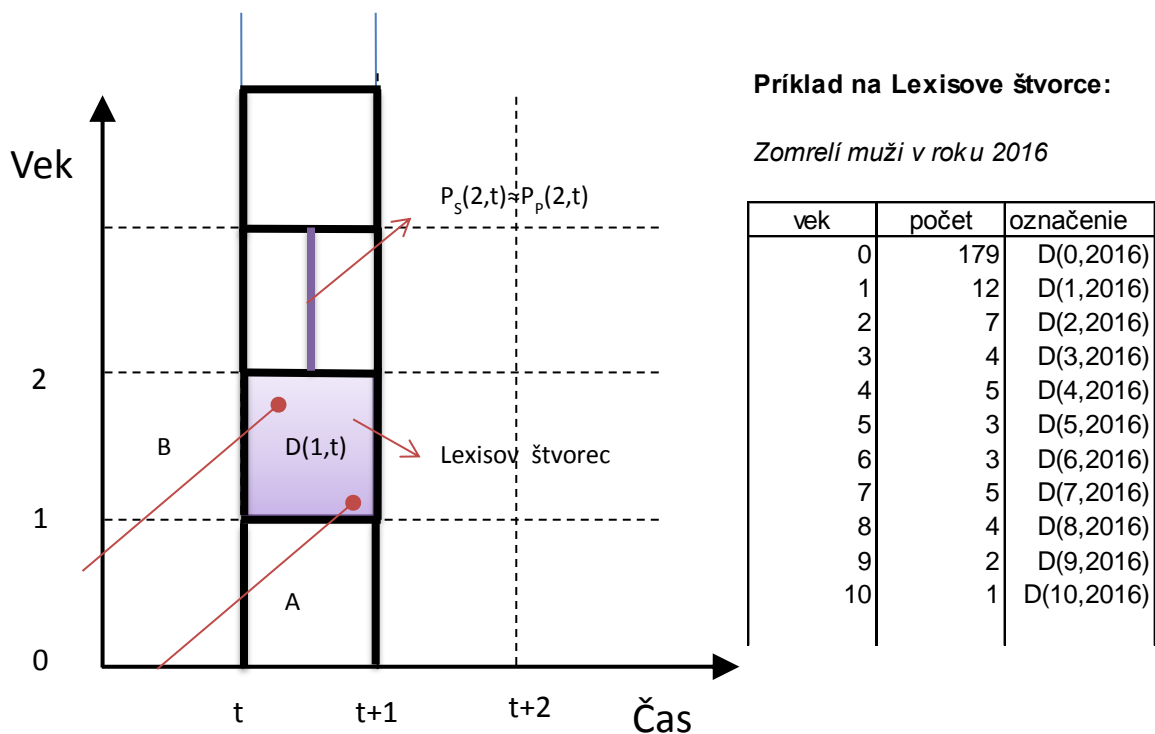
Úvod

Úmrtnostné tabuľky (ďalej UT) sú najviac rozpracované nástroje na meranie úmrtnostných pomerov určitej populácie. Sú založené na pravdepodobnosti úmrtia resp. prežitia, na základe ktorých sú vypočítané dĺžky života populácie. Existujú rôzne typy UT a sú popísané v úvode [4]. V tomto materiáli sa budeme zaoberať **generačnou úmrtnostnou tabuľkou**, ktorá v slovenskej praxi ešte nebola použitá.

Definície a použité symboly

Generácia¹ je skupina osôb, ktorá sa zúčastnila určitej demografickej udalosti v rovnakom časovom intervale. Môže to byť generácia zosobášených, alebo rozvedených, alebo narodených, ktorý sa zosobášili, alebo rozvedli, alebo sa narodili rovnakom časovom intervale, čo môže byť napríklad jeden určitý rok. My sa budeme zaoberať **generáciami narodených** v jednotlivých rokoch. Budeme sledovať vymieranie jednotlivých generácií pomocou úmrtnostných tabuliek a vytvárať rôzne charakteristiky súvisiace s týmto vymieraním. Budeme postupovať na základe metodiky, ktorú publikovali autori „Human Mortality Database HMD“ v materiáli [3]. Vytvoríme UT na základe zomrelých triedených do **Lexisových trojuholníkov** a populácie k 31.12. každého roku skúmania.

Obr. 1.: Lexisov diagram organizácie dát pre prierezové UT za kalendárny rok t



¹ Generácia sa môže nazývať aj kohorta, sú to synonymá. Niektorí autori používajú slovo generácia, ak sa dá populácia geograficky ohraničiť a slovo kohorta, ak sa nedá. Napr. kohorta Rómov, generácia populácie Slovenska.

Označenie:

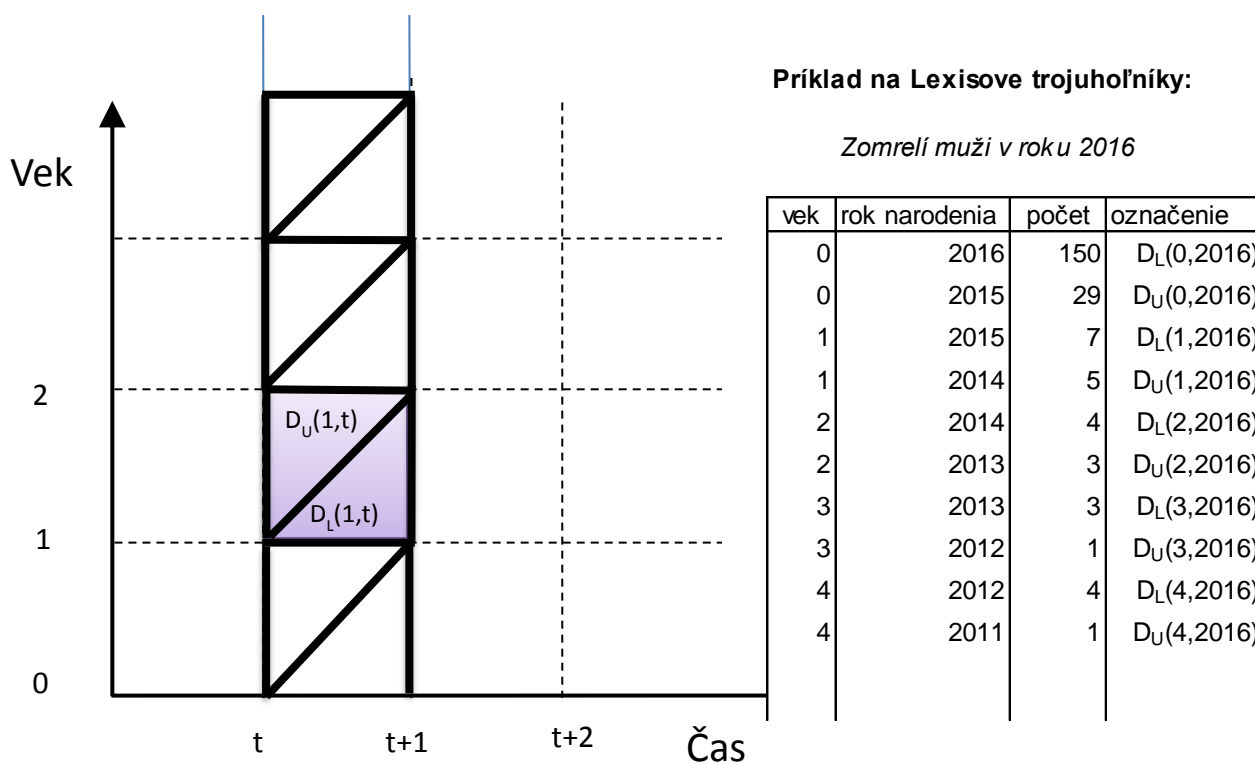
$P_S(x+1,t)$ populácia k 1.7. roku t vo veku x+1

$P_P(x+1,t)$ priemerný stav populácie roku t vo veku x

$D(x,t)$ počet zomrelých v roku t vo veku x (tzv. Lexisov štvorec)

Ako z Obr. 1 vidno, pri prierezovom pohľade sme pracovali z údajmi triedené za jeden kalendárny rok a dosiahnutého veku. V Lexisovom štvorci $D(x,t)$ sú spočítané zomrelí, ktorí zomreli vo veku x a v roku t. Čiary života A a B naznačujú, že sú tu osoby, ktoré patria do dvoch generácií, a to osoby typu A sa narodili v roku t-1 a osoby typu B sa narodili v roku t-2. Sú to dve generácie, ktoré chceme osobitne skúmať, a tak potrebujeme Lexisové štvorce rozdeliť na Lexisové trojuholníky (Obr.2.)

Obr.2.: Rozdelenie Lexisových štvorcov na Lexisové trojuholníky za kalendárny rok t



Označenie:

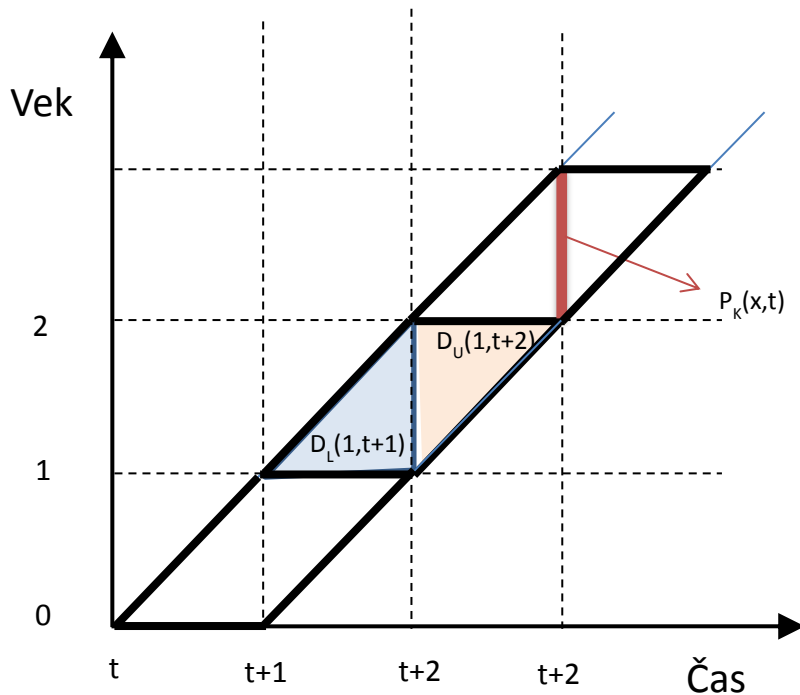
$D_L(x,t)$ počet zomrelých v roku t vo veku x a narodených v roku t-x (tzv. Lexisov dolný trojuholník)

$D_U(x,t)$ počet zomrelých v roku t vo veku x a narodených v roku t-x (tzv. Lexisov horný trojuholník)

Platí: $D(x, t) = D_L(x, t) + D_U(x, t)$

Vzhľadom na to, že naším cieľom je skúmať vymieranie konkrétnej generácie, musíme tabuľky zomrelých podľa Lexisových trojuholníkov za kalendárne roky preskupiť podľa generácii, ako to je zobrazené na Obr.3.

Obr.3. Lexisové trojuholníky podľa generácie t



Označenie:

$P_k(x, t)$ populácia k 31.12. roku t vo veku x

$D_L(x, t+1)$ počet zomrelých v roku t+1 vo veku x a narodených v roku t-x (tzv. Lexisov dolný trojuholník)

$D_U(x, t+2)$ počet zomrelých v roku t+2 vo veku x a narodených v roku t-x (tzv. Lexisov horný trojuholník)

Príklad na Lexisové trojuholníky podľa generácie 1948 mužov

vek	rok úmrtia		rok narodenia/ generácia	zomrelí v D_L	zomrelí v D_U
	pre dolný trojuhoľník D_L	pre horný trojuhoľník D_U			
0	1948	1949	1948	4503	1537
1	1949	1950	1948	322	146
2	1950	1951	1948	90	79
3	1951	1952	1948	37	34
4	1952	1953	1948	31	35
5	1953	1954	1948	15	25

Premenné, ktoré budú vytvorené v UT pre narodených v roku t (generácia t):

$m_t(x)$	špecifická miera úmrtnosti medzi vekmi x a x+1 pre narodených v roku t
$q_t(x)$	pravdepodobnosť úmrtia medzi vekmi x a x+1 pre narodených v roku t
$p_t(x)$	pravdepodobnosť prežitia medzi vekmi x a x+1 pre narodených v roku t
$d_t(x)$	tabuľkový počet zomrelých medzi vekmi x a x+1 pre narodených v roku t
$l_t(x)$	počet prežívajúcich v presnom veku x z $l_t(0)=100000$ pre narodených v roku t
$L_t(x)$	počet človekorokov prežitých medzi vekmi x a x+1 pre narodených v roku t
$T_t(x)$	počet človekorokov zostávajúcich po presnom veku x pre narodených v roku t
$e_t(x)$	zostávajúca stredná dĺžka života v presnom veku x v rokoch pre narodených v roku t

Vzťahy medzi premennými:

$$m_t(x) = \frac{D_L(x, t) + D_U(x, t + 1)}{P_K(x, t) + \frac{1}{3}(D_L(x, t) - D(x, t + 1))}$$

$$q_t(x) = \frac{D_L(x, t) + D_U(x, t + 1)}{P_K(x, t) + D_L(x, t)}$$

$$p_t(x) = 1 - q_t(x)$$

$$l_t(1) = l_t(0) \cdot p_t(0)$$

$$l_t(x) = l_t(0) \prod_{i=0}^{x-1} p_t(i)$$

$$d_t(x) = l_t(x) \cdot q_t(x)$$

$$L_t(0) = l_t(0) - 0,9 \cdot d_t(0)$$

$$L_t(x) = l_t(x) - 0,5 \cdot d_t(x) \quad \text{pre } x > 0$$

$$T_t(x) = \sum_{i=x}^{\omega} L_t(i) \quad \text{kde } \omega \text{ je najvyšší vek}$$

$$e_t(x) = \frac{T_t(x)}{l_t(x)}$$

Vek ω sme zvolili vek 100 rokov, vzhľadom na to, že vekové štruktúry obyvateľov sú len do tohto veku.

Použitá databáza

Pri tvorbe databázy pre generačné UT budeme potrebovať dáta D_L , D_U a P_K . Vychádzať budeme z dát, ktoré boli publikované Štatistickým úradom SR, Federálnym štatistickým úradom Československa a Štatistickým úradom Českej republiky. Zistili sme, že reálne by sme mohli získať dáta o úmrtí od roku 1919 prierezovom pohľade a dáta o žijúcej populácie od roku 1920 určitými nedostatkami, ktoré budeme musieť určitými úpravami doplniť.

Na získanie údajov o úmrtí boli použité knižné publikácie „Pohyb obyvateľstva...“ vydané príslušným štatistickým úradom, či už za ČSSR, ČSFR, SSR alebo SR. Za roky od 1925 až 1937 a od 1945 po 2016 sme získali potrebné dáta triedené podľa pohlavia, veku, roku výskytu a roku narodenia (Lexisové trojuholníky D_L a D_U).

Za roky 1919 až 1924 sme z knižných publikácií „Pohyb obyvateľstva...“ získali počty zomrelých podľa pohlavia a veku (Lexisové štvorce D). Chýbala informácia o roku narodenia. Na rozčlenenie údajov podľa veku a roku narodenia, čiže do D_L a D_U sme použili metódu publikovanú v [3] (Methods Protocol). Metóda (označená v [3] ako Model I) sa dá vyjadriť matematicky nasledovne:

$$D_L(x, t) = \text{koef}(x) \cdot D(x, t)$$

$$D_U(x, t) = (1 - \text{koef}(x)) \cdot D(x, t)$$

$$\text{koef}(x) = \text{konst} + \text{vplyv_veku}(x) \quad , \text{ kde}$$

	Muži	Ženy		Muži	Ženy
konst	0,5226	0,5170			
vplyv_veku			vplyv_veku		
0	0,2294	0,2285	50-54	-0,0201	-0,0136
1	0,0370	0,0462	55-59	-0,0195	-0,0140
2-4	-0,0440	-0,0003	60-64	-0,0206	-0,0180
5-9	-0,0088	-0,0101	65-69	-0,0221	-0,0215
10-14	-0,0200	-0,0254	70-74	-0,0236	-0,0233
15-19	-0,0274	-0,0242	75-79	-0,0255	-0,0251
20-24	-0,0056	-0,0162	80-84	-0,0240	-0,0235
25-29	-0,0014	-0,0141	85-89	-0,0169	-0,0165
30-34	-0,0056	-0,0127	90-94	-0,0077	-0,0066
35-39	-0,0145	-0,0153	95-99	0,0063	0,0055
40-44	-0,0189	-0,0142	100-104	0,0348	0,0274
45-49	-0,0210	-0,0131			

Najmenej informácii sme získali o úmrtí za roky 1938 až 1944. Z publikácie [1] sme získali len celkové počty zomrelých za jednotlivé roky bez informácií o potrebnej štruktúre, či už pohlavnej alebo vekovej. Na rozčlenenie celkových počtov do pohlavnej štruktúry sme použili koeficienty maskulinity a femininity (podiel mužov resp. žien k celkovej populácie) pre jednotlivé roky. Keďže sme nenašli relevantné publikované metódy na rozloženie týchto čísel do jednotlivých vekov, použili sme postup, ktorý sa zakladá na vekových štruktúrach

zomrelých za roky, kde je táto štruktúra známa. Východiskové štruktúry sme zvolili za roky 1937 a 1945. Vytvorili sme umelú vekovú štruktúru spriemerovaním týchto dvoch štruktúr zvlášť pre mužov a zvlášť pre ženy. Na základe týchto štruktúr sme potom vytvorili vekové rozloženie úmrtí za roky 1938 až 1944 v tvare D. Ešte bolo potrebné rozčleniť túto štruktúru do tvaru D_L a D_U , čo sme vykonali metódou publikovanú v [3] (Methods Protocol).

Na získanie údajov o počte obyvateľov k 31.12. (P_K) boli použité knižné publikácie „Vekové zloženie obyvateľstva...“ za jednotlivé roky, vydané príslušným štatistickým úradom, či už za ČSSR, ČSFR, SSR alebo SR a publikáciu [5]. Počet obyvateľov k 31.12 sme podľa veku a pohlavia sme získali za roky 1950 až 2016.

Za ostatné roky, a to 1920 až 1949, sme získali sumárne hodnoty počtu obyvateľov podľa pohlavia k 31.12 z publikácie [1] a bolo potrebné tieto hodnoty rozčleniť do vekovej štruktúry podľa pohlavia.

Použili sme na to štruktúru obyvateľov k 1.7. podľa pohlavia a veku (P_S). Tieto počty sú publikované za roky od 1920 až 1938 a od roku 1945 doteraz. Za chýbajúce roky od 1939 až 1944 máme k dispozícii iba sumárne počty. Podobne ako pri počtoch umretých sme tieto súčty rozčlenili do jednotlivých vekov pomocou umelej štruktúry, ktorú sme dostali spriemerovaním štruktúr za roky 1937 a 1945 podľa pohlavia. Takýmto spôsobom sme získali kompletne vekové štruktúry obyvateľov k 1.7. za celé obdobie 1920 až 2016.

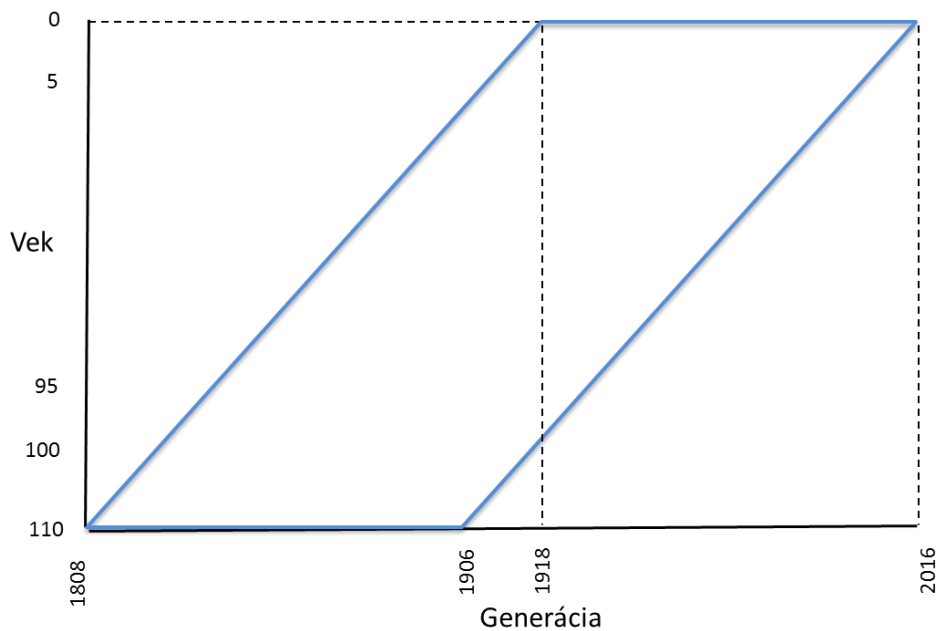
Chýbajúce vekové štruktúry obyvateľov k 31.12. sme dostali použitím vekových štruktúr k 1.7 za roky 1920 až 1949. Urobili sme to tak, že známe sumárne hodnoty k 31.12 sme rozčlenili do vekovej štruktúry proporcionálne podľa štruktúry k 1.7.

Na základe takto získaných D_L , D_U a P_K môžeme vytvoriť generačné úmrtnostné tabuľky pre Slovensko pre mužov a pre ženy. Treba tu však poznamenať, že premenné UT pre jednotlivé veku, ktoré sme popísali vyššie, je možné vytvoriť len vtedy, ak sú známe všetky tri hodnoty D_L , D_U a P_K pre daný vek. Vzhľadom na to, že dáta od roku 1920 nepokrývajú ani jednu generáciu kompletne, výsledné UT budú nekompletné. Budú chýbať parametre pre niektoré veku, ktoré však v nasledujúcich rokoch bude možné na základe nových údajov dopĺňať.

Vyhodnotenie

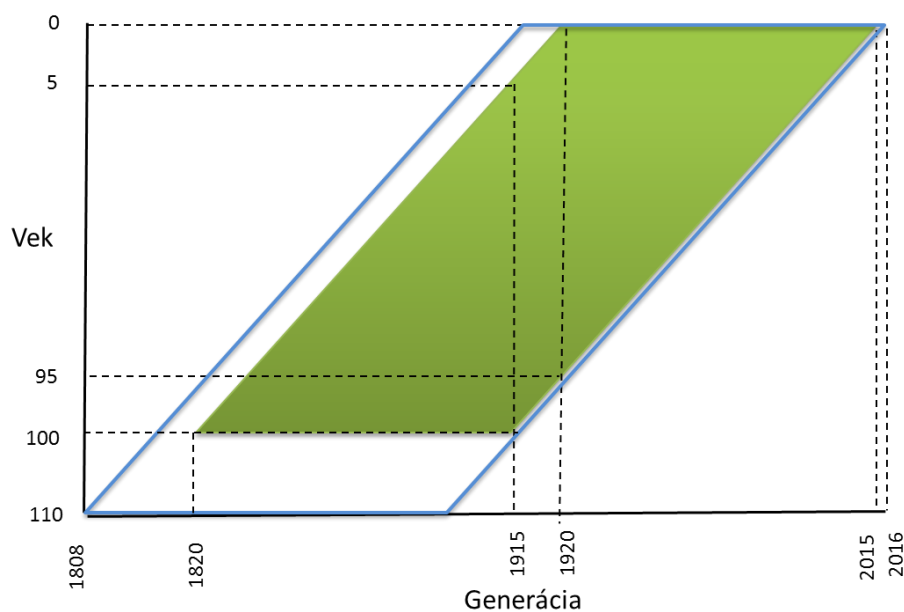
Z dostupných údajov do roku 2016, popísaných vyššie, sme vytvorili generačné tabuľky zomrelých pre mužov a pre ženy informáciami od roku 1808 po 2016. Z generácie 1808 máme len jednu známu hodnotu, a to $D_U = 0$ pre mužov a aj pre ženy, t.z. 0 osôb zomrelo z tejto generácie v roku 1919 vo veku 110 rokov. Z ďalších generácií máme postupne viac informácií. Graficky môžeme tieto informácie zobraziť generačným – vekovým priestorom zobrazeným nasledujúcim grafom. Dostupné údaje zomrelých sú zobrazené priestorom ohraničeným modrou čiarou.

Graf 4. Generačno-vekový priestor dostupnosti dát tabuľky zomrelých pre Slovensko v roku 2016



Ako z grafu vidno prvé známe hodnoty zomrelých 0 ročných sú pre generáciu 1918, posledné z roku 2016. Pre zomrelých 110 ročných prvé známe hodnoty sú z roku 1808 a posledné z roku 1906.

Graf 4. Generačno-vekový priestor dostupnosti dát pre pravdepodobnosti úmrtia pre Slovensko v roku 2016

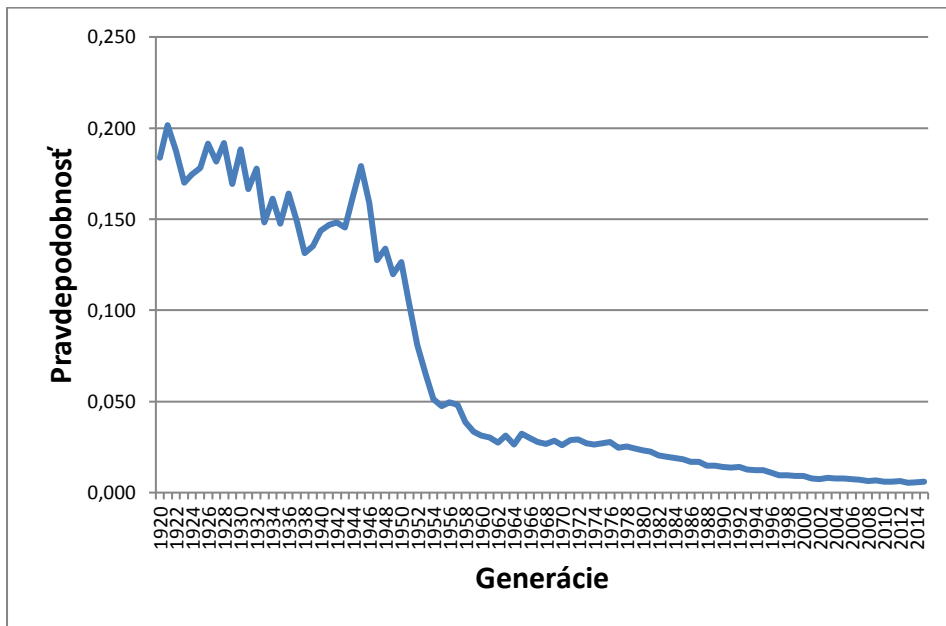


Prvým krokom na vytvorenie generačných úmrtnostných tabuliek je výpočet pravdepodobnosti úmrtia z veku x na vek $x+1$. K tomu je potrebné poznať všetky tri hodnoty

D_L , D_U a P_K . Vzhľadom na to, že veľkosť žijúcej populácie P_K sú známe len do veku 100 rokov, celý dostupný generačno-vekový priestor musíme skrátiť do tohto veku. Aj niektoré generácie musíme vynechať u ktorých nie sú známe všetky tri uvedené hodnoty. Takže nakoniec dostaneme priestor vyznačený na grafe zelenou farbou.

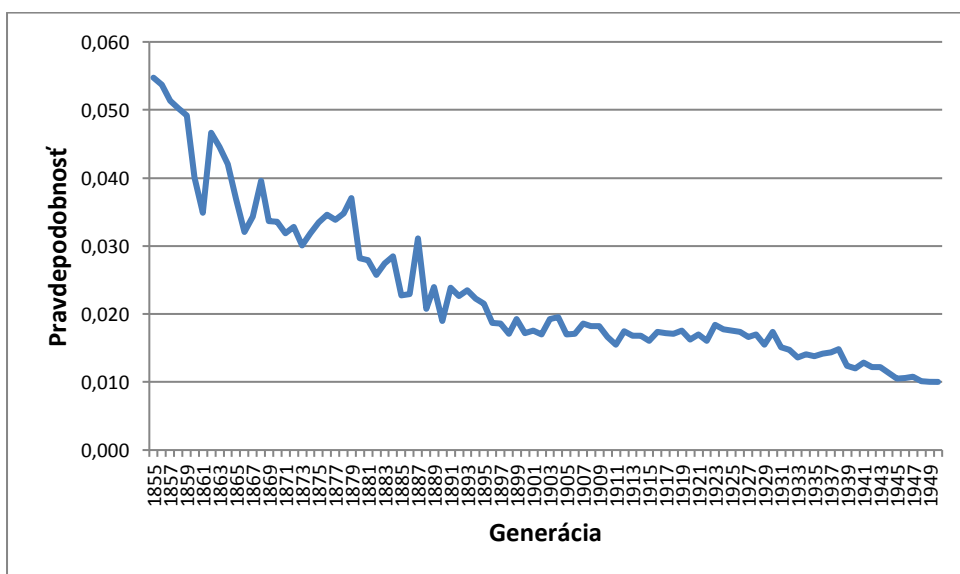
Ako príklad z tabuľky pravdepodobnosti úmrtia pre generácie môžeme uviesť vývoj hodnôt pre 0 ročných mužov pre generácie 1920 až 2015.

Graf 5. Generačné pravdepodobnosti úmrtia 0 ročných mužov na Slovensku



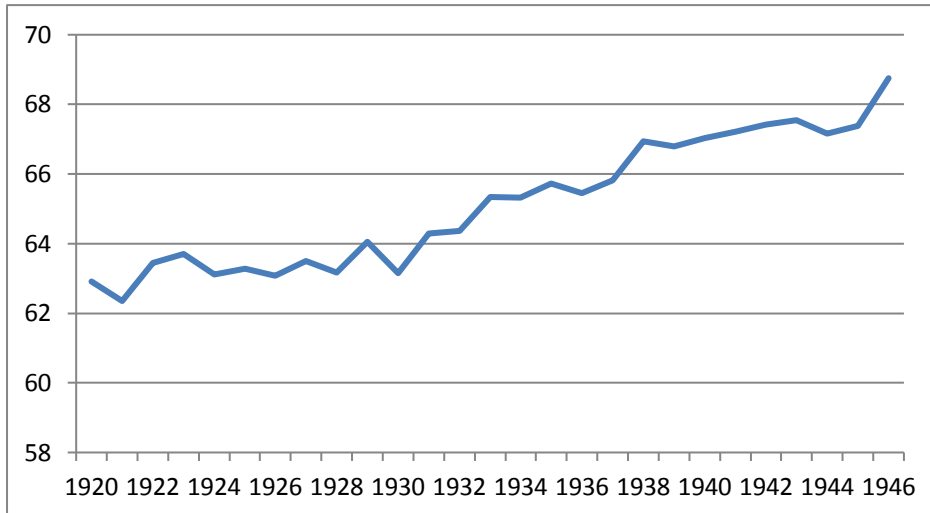
Ako z grafu vyplýva, muži generácie 1920 až 1945 mali výrazne vyššiu pravdepodobnosť, že sa nedožijú prvého veku života ako generácie 1960 až 2015.

Graf 5. Generačné pravdepodobnosti úmrtia 65 ročných žien na Slovensku



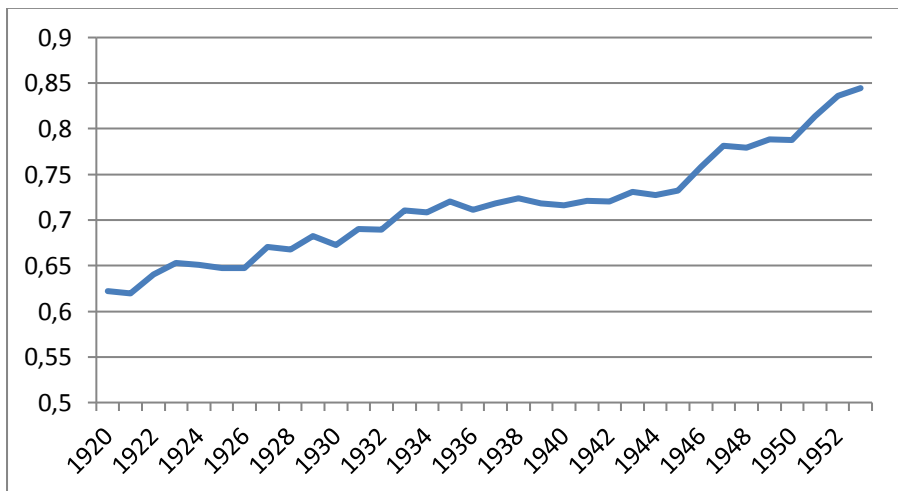
Ďalší príklad z tabuľky pravdepodobnosti úmrtia pre generácie uvádzame pravdepodobnosti úmrtia 65 ročných žien. Ako vidno máme k dispozícii iný rozsah generácii ako pri 0 ročných, a to pre generácie od 1855 až 1949.

Graf 6. Mediánový vek tabuľkových zomrelých mužov na Slovensku



Mediánový vek zomrelých mužov, čo znamená, vek v ktorom zomrie polovica tabuľkových narodených (100000 narodených) je možné vypočítať pre generácie 1920 až 1946. U ďalších generácií ešte nevymrelo polovica tabuľkových narodených. Rastúci graf naznačuje posun ťažiska úmrtí do vyšších rokov.

Graf 7. Pravdepodobnosť dožitia sa veku 62 rokov u žien na Slovensku

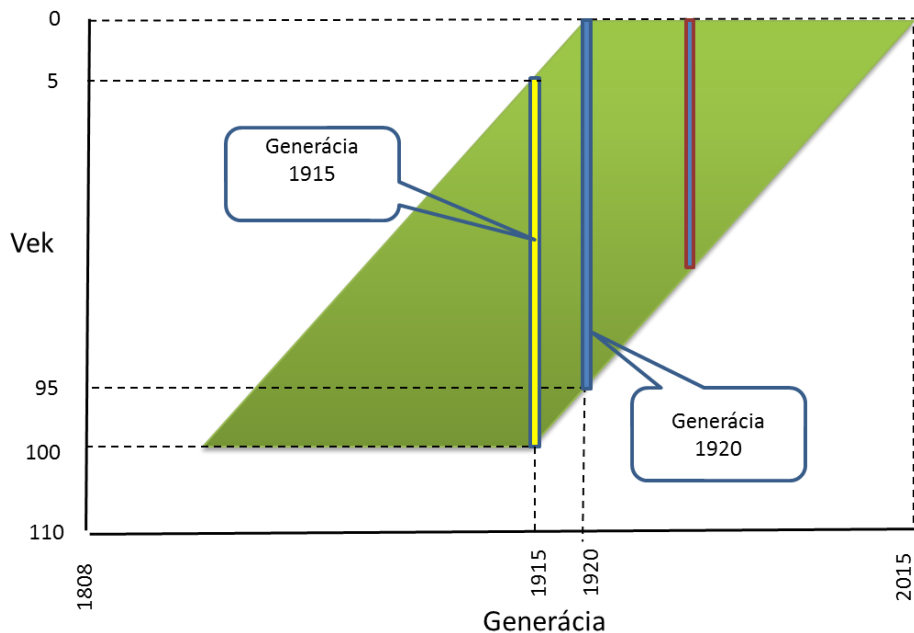


Graf pravdepodobnosti dožitia sa veku 62 rokov (ukazovateľ sa nazýva aj ako kumulatívna pravdepodobnosť) ukazuje rastúcu šancu dožiť sa dôchodkového veku.

Naším hlavným cieľom je vytvoriť generačné úmrtnostné tabuľky. Ako s algoritmu popísaného úvode tohto materiálu vyplýva, k tomu potrebujeme pravdepodobnosti úmrtia pre celú škálu vekovej štruktúry, a to od veku 0 rokov až po vek 100 rokov. Ako

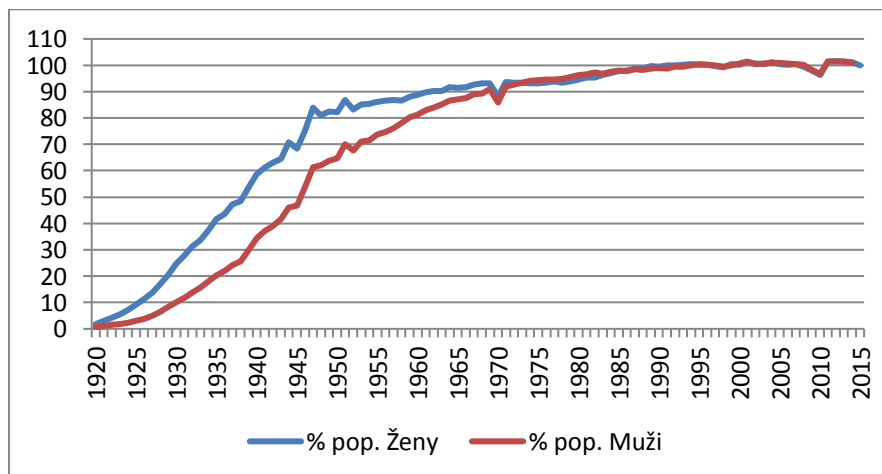
z nasledujúceho grafu vyplýva zatiaľ nemáme generáciu, ktorá by takéto požiadavky splnila. Najviac naplnené generácie sú od 1915 až 1920. Avšak generácie 1919 a staršie nemajú informáciu o veku 0, bez ktorej sa úmrtnostná tabuľka nedá vytvoriť. Takže prvá použiteľná generácia je 1920, v ktorej máme údaje za veku 0 až 95 rokov. Tu môžeme vytvoriť neúplnú úmrtnostnú tabuľku, čo bude reprezentovať úmrtnostné pomery generácie do veku 95 rokov. Ak získame pomery úmrtia u mužov pre 319 osôb, u žien pre 863 osôb starších ako 95 rokov môžeme tabuľku skompletizovať.

Graf 8. Vytvorené generačné úmrtnostné tabuľky na Slovensko v roku 2016



Je zrejmé, že chýbajúce pomery úmrtia u mladších generácií rastie, a tak ich úmrtnostné tabuľky viac neúplné a dávajú skreslené informácie. Neúplnosť môžeme prezentovať nasledujúcim grafom, čo obsahuje percento ešte žijúcej populácie pre jednotlivé generácie.

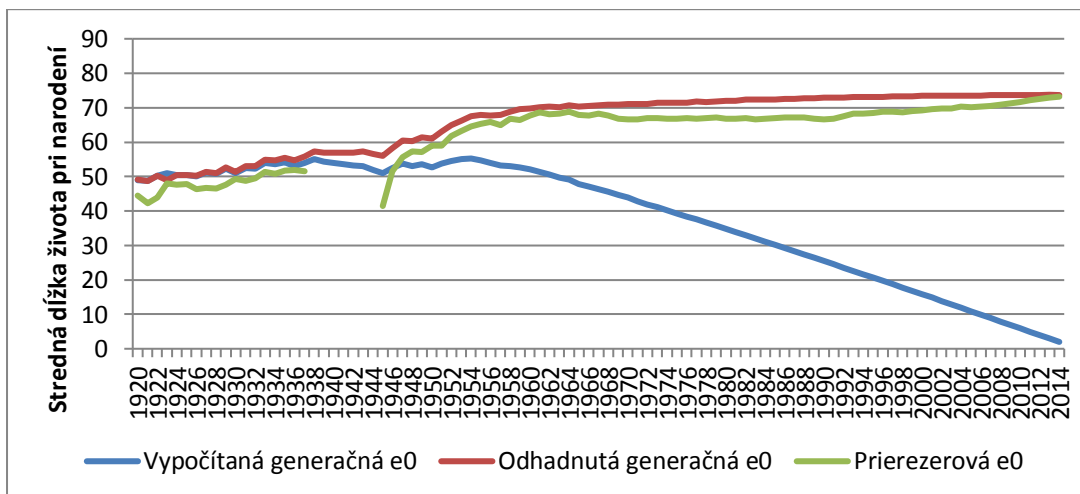
Graf 9. Percento žijúcej populácie pre generácie na Slovensku v roku 2016



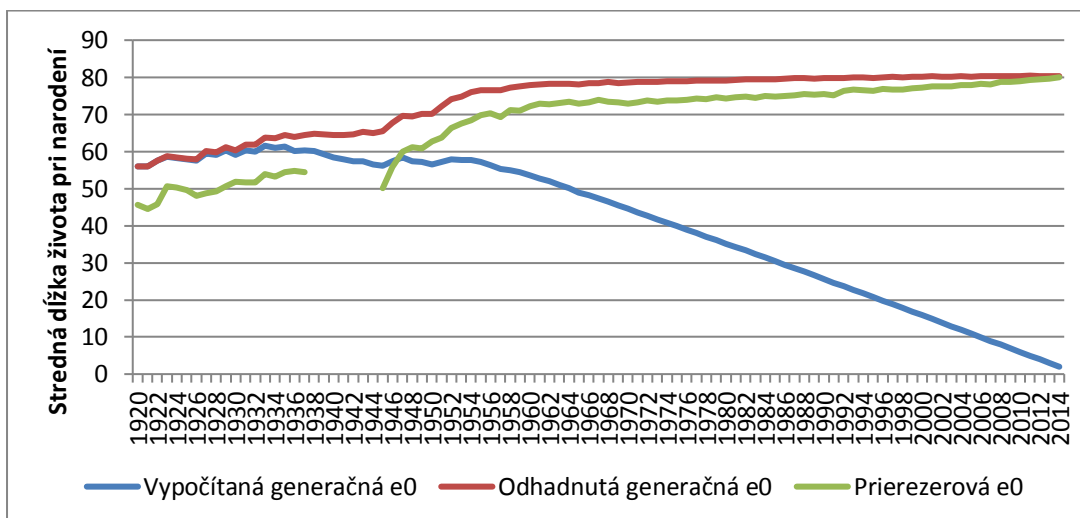
Aby boli takto skonštruované generačné úmrtnostné tabuľky použiteľné, resp. dávali presnejšie výsledky, je možné tabuľky doplniť odhadnutými informáciami. Je potrebné odhadnúť pravdepodobnosti úmrtia pre chýbajúce veku. Je tu viac možností, napr. predĺžiť pravdepodobnosti úmrtia extrapoláciou pomocou lineárnej funkcie, alebo pomocou funkcie, ktoré sa používajú pri prierezových úmrtnostných tabuľkách [4], alebo použitie pravdepodobnosti z prierezových úmrtnostných tabuliek za posledný rok. Výber metódy a získané odhady chýbajúcich hodnôt však je potrebné podrobiť analýze. Zistiť vierohodnosť výsledkov, ktorá metóda dáva výsledky, ktoré nie sú veľmi skreslené umelým zásahom. Táto analýza však presahuje rámec tejto práce a je potrebné sa tomu venovať v budúcnosti.

Ako príklad uvádzame stredné dĺžky života pri narodení z vytvorených generačných tabuliek vypočítaných bez extrapolácie chýbajúcich hodnôt, odhadnutých pomocou pravdepodobností úmrtia za rok 2016 a prierezových úmrtnostných tabuliek pre mužov a žien.

Graf 10. Stredná dĺžka života pri narodení mužov na Slovensku



Graf 10. Stredná dĺžka života pri narodení žien na Slovensku



Literatúra

- [1] Demografická příručka. Český statistický úřad. Praha 1996.
- [2] Demografická příručka. Federální statistický úřad. Praha 1982.
- [3] *Human Mortality Database*. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). (www.mortality.org alebo www.humanmortality.de)
- [4] Mészáros, J.: Požiadavky modernizácie demografickej štatistiky vo vzťahu k novým metódam graduácie pravdepodobností úmrtia. Pracovný materiál. INFOSTAT Bratislava, 2016. (www.infostat.sk/vdc/pdf/TeoriaUT.pdf)
- [5] Věkové složení obyvatelstva v letech 1920 -1937 a 1945 – 1979 (ČSSR, ČSR, SSR), Český statistický úřad