

Zhodnotenie metodických postupov konštrukcie prierezových a generačných úmrtnostných tabuliek a identifikácia hlavných rozdielov

Autor: Ján Mészáros

Analyticko-metodologický materiál

Číslo čiastkovej úlohy 900-12/1

V demografickej praxi na Slovensku na skúmanie úmrtnosti sa doteraz používali **prierezové úmrtnostné tabuľky**. Ich popis tvorby je zdokumentovaný v publikácii [3] a je aplikovaný na dáta od roku 1920. Ich použitie je spojená so syntetickou skupinou ľudí, ktorí zažívajú hypoteticky demografické podmienky z krátkeho časového intervalu počas celého ich života. Napr. na základe mier úmrtnosti podľa jednotlivých vekov v danom období sa konštruje snímok života hypotetickej populácie naraz narodených. Rozsah tejto hypotetickej populácie sa volá **koreň** a volí sa obvykle 100 000 ľudí. Dá sa vypočítať teoretický priemerný vek pri úmrtí tejto hypotetickej populácie, ktorí sú vystavení vplyvu faktorov úmrtností z krátkeho časového intervalu, napr. jedného kalendárneho roku, počas ich celého života.

Generačné úmrtnostné tabuľky, na rozdiel od prierezových, sú spojené so skutočnými generáciami narodených v krátkom časovom intervale, spravidla jeden rok, a zažívajú skutočné demografické podmienky z ich celého života. Predstavujú záznam priebehu života konkrétnej populácie súčasne narodených jedincov zahájený okamžikom narodenia všetkých jedincov tejto populácie a končiaci úmrtím posledného z nich.

Generačné úmrtnostné tabuľky Slovenskej republiky boli vytvorené v roku 2017. Ich tvorba a časové pokrytie bolo podriadené dostupnými potrebnými dátami. Použili sme dáta publikované Štatistickým úradom Slovenskej republiky, Federálnym štatistickým úradom Československa a Štatistickým úradom Českej republiky. Získali sme dáta o úmrtí od roku 1919 prierezovom pohľade a dáta o žijúcej populácie od roku 1920. Vzhľadom na to, že nie všetky údaje o úmrtí neboli v požadovanej štruktúre, čiže počty úmrtí neboli triedené podľa veku a roku narodenia, museli sme vykonať určité matematické úpravy na nich.

Tabuľka 1.: Dostupné dáta o úmrtí za Slovenskú republiku k roku 2016

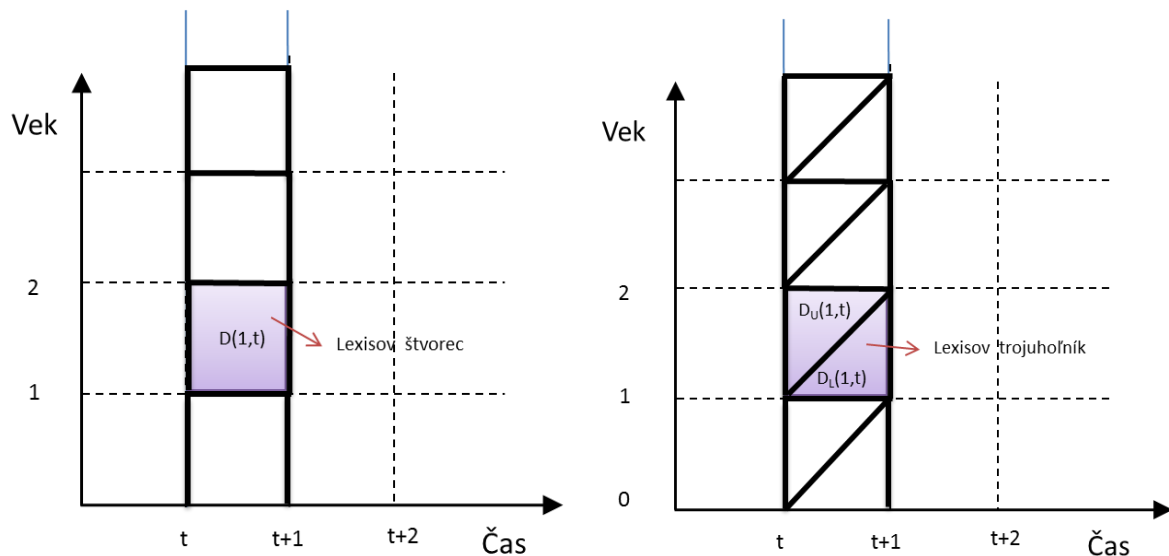
vek/rok	1919	-	1924	1925	-	1937	1938	-	1944	1945	-	2016
0												
0												
1												
1												
2												
2												
3												
3												
...
...
Súčet												

V tabuľke 1 sme zobrazili dostupné dáta o úmrtí a naznačili potrebné úpravy. So žltou farbou sme označili roky za ktoré máme kompletne potrebné dáta a nebolo potrebné vykonať žiadne zmeny. Sú to roky 1925 až 1937 a 1945 až 2016, kde sme získali počty zomrelých triedených podľa pohlavia, veku a roku narodenia.

Za obdobie 1919 až 1924, sú označené zelenou farbou, boli k dispozícii počty zomrelých podľa pohlavia a veku ale bez rozdelenia podľa roku narodenia. Na rozloženie počtu zomrelých podľa roku narodenia sme použili algoritmus popísaný v metodickom materiáli [4]. Prakticky tu ide,

podľa demografickej terminológie o prechod z „Lexisových štvorčiek“ na „Lexisové trojuholníky“ v prierezovom pohľade.

Graf 1. Príklad zobrazenia „Lexisových štvorčiek a trojuholníkov“ v prierezovom pohľade



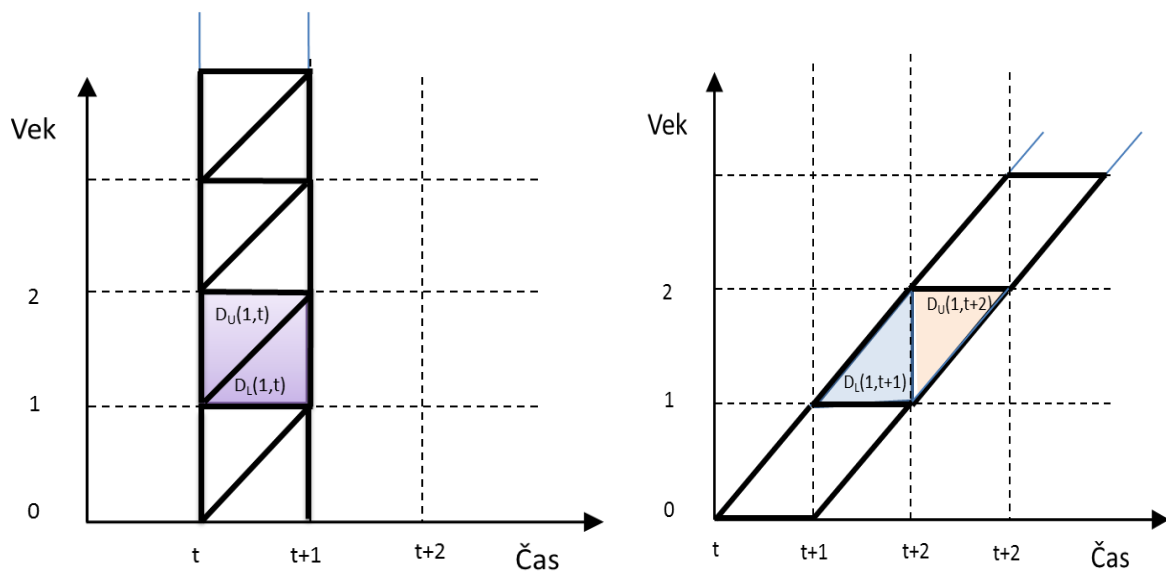
Tabuľka 2. Príklad tabuliek s „Lexisovými štvorčkami a trojuholníkmi“ prierezovom pohľade

Lexisove štvorce:			Lexisove trojuholníky:			
<i>Zomrelí muži v roku 2016</i>			<i>Zomrelí muži v roku 2016</i>			
vek	počet	označenie	vek	rok narodenia	počet	označenie
0	179	D(0,2016)	0	2016	150	D _L (0,2016)
1	12	D(1,2016)	0	2015	29	D _U (0,2016)
2	7	D(2,2016)	1	2015	7	D _L (1,2016)
3	4	D(3,2016)	1	2014	5	D _U (1,2016)
4	5	D(4,2016)	2	2014	4	D _L (2,2016)
5	0	D(5,2016)	2	2013	3	D _U (2,2016)
6	0	D(6,2016)	3	2013	3	D _L (3,2016)
7	0	D(7,2016)	3	2012	1	D _U (3,2016)
8	0	D(8,2016)	4	2012	4	D _L (4,2016)
9	0	D(9,2016)	4	2011	1	D _U (4,2016)
10	0	D(10,2016)				

Za obdobie 1938 až 1944 sme získali len sumárne údaje o zomrelých, označené modrou farbou v tabuľke 1. Tieto sumáre bolo potrebné rozhodnúť do štruktúry podľa veku a roku narodenia. Vykonali sme to pomocou umelej štruktúry vytvorenej spríemerovaním štruktúry za rok 1937 a 1945.

Po vykonaní uvedených prepočtov sme ešte museli pristúpiť k preskupeniu „Lexisových trojuholníkov“ z prierezového pohľadu na generačný pohľad. Grafické znázornenie je na grafe 3.

Graf 2. Príklad preskupenia z prierezového pohľadu na generačný pohľad



Tabuľka 3. Príklad tabuliek s „Lexisovými trojuholníkmi“ generačnom pohľade (generácia 1948)

vek	rok úmrtia		rok narodenia/ generácia	zomrelí v D_L	zomrelí v D_U
	pre dolný trojuholník D_L	pre horný trojuholník 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

Pre tvorbu generačných úmrtnostných tabuliek potrebujeme aj počty žijúcich osôb k 31.12. pre obdobie od 1920 po 2016 triedený v každom roku podľa veku.

Ako z tabuľky 4 vyplýva, kompletne vyhovujúce údaje k 31.12 (koncové stavy) sme získali od roku 1950. Za roky 1920 až 1937 a 1945 až 1949 sme získali len sumárne údaje za koncové stavy a štruktúru populácie k 1.7. (stredné stavy). Sumárne údaje k 31.12. sme rozpočítali podľa štruktúry 1.7. Ostávali nám vyriešiť roky 1938 až 1944 pre, ktoré sme zistili iba sumárne hodnoty k 31.12. Podobne ako pri počtoch zomrelých sme tieto hodnoty rozhodili na základe umelej štruktúry, ktorá vznikla spriemerovaním štruktúr za roky 1937 a 1945.

Tabuľka 4.: Dostupné dáta o populácii za Slovenskú republiku k roku 2016

vek/rok	1920	-	1937	1938	-	1944	1945	-	1949	1950	-	2016
0	1.7	1.7	1.7				1.7	1.7	1.7	31.12	31.12	31.12
1	1.7	1.7	1.7				1.7	1.7	1.7	31.12	31.12	31.12
2	1.7	1.7	1.7				1.7	1.7	1.7	31.12	31.12	31.12
3	1.7	1.7	1.7				1.7	1.7	1.7	31.12	31.12	31.12
...
...
Súčet	31.12	31.12	31.12	31.12	31.12	31.12	31.12	31.12	31.12	31.12	31.12	31.12

Takto získané údaje za žijúcu populáciu k 31.12. sme priradili do tabuliek zomrelých v generačnom pohľade spárovaných podľa veku a roku úmrtia pre trojuholník D_L . Takto vznikli tabuľky vstupných dát pre tvorbu generačných úmrtnostných tabuliek.

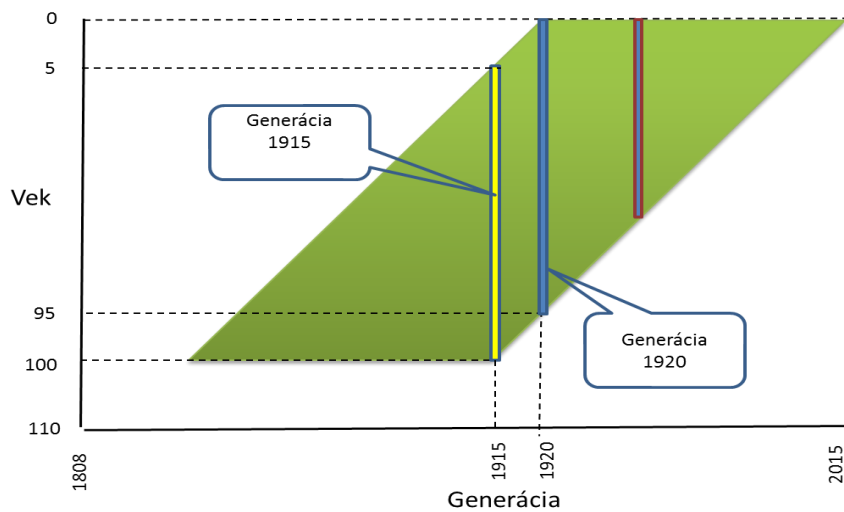
Tabuľka 5. Príklad vstupných dát pre tvorbu generačnej úmrtnostnej tabuľky pre generáciu 1920

vek	rok úmrtia		rok narodenia	zomrelí		populácia
	pre D_L	pre D_U	generácia	D_L	D_U	31.12 pre L
0	1920	1921	1920	7625	2899	49679
1	1921	1922	1920	1266	764	39386
2	1922	1923	1920	392	319	38086
3	1923	1924	1920	172	183	37594
4	1924	1925	1920	120	94	37239
5	1925	1926	1920	82	80	37017
6	1926	1927	1920	90	88	36838
7	1927	1928	1920	70	65	36657
8	1928	1929	1920	67	55	36519
9	1929	1930	1920	56	46	36388

Na vypočítanie všetkých riadkov a stĺpcov generačných UT sú potrebné známe hodnoty zomrelých D_L , D_U a populácie k 31.12. pre všetky veky. Prvým krokom je vypočítanie miery úmrtnosti a pravdepodobnosti úmrtia.

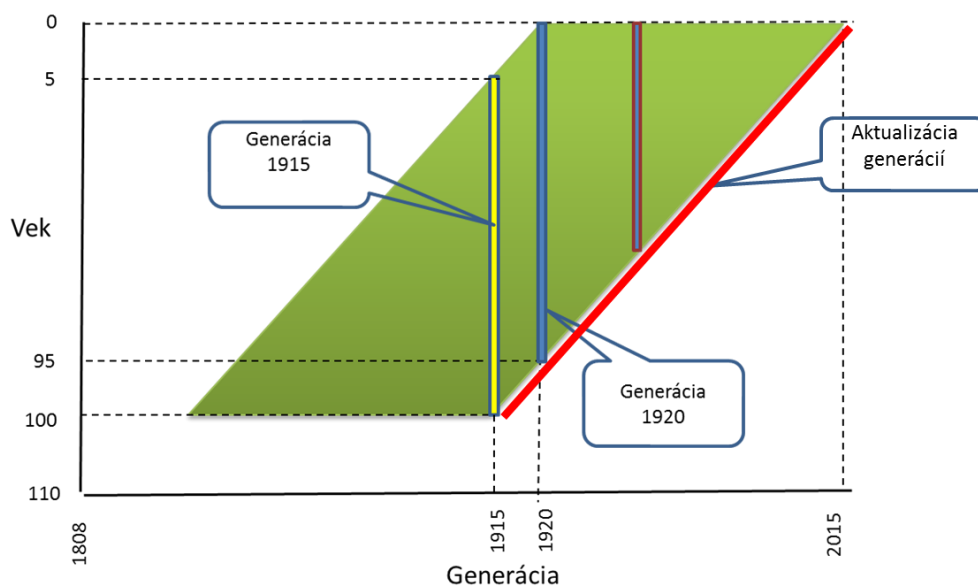
Ak niektorá, z uvedených troch hodnôt chýba, nie je možné miery úmrtnosti a pravdepodobnosť úmrtia vypočítať. **V slovenských podmienkach nemáme zatiaľ generáciu, ktorá by mala miery úmrtnosti a pravdepodobnosti úmrtia vypočítané pre všetky veky generácie.** Sú generácie, ktorým chýbajú údaje pre mladšie veky (napr. generácia 1915) a sú generácie kde chýbajú staršie veky (napr. generácia 1920). Úmrtnostné tabuľky generácií, ktoré vzniknú z týchto vypočítaných pravdepodobností úmrtia nie sú kompletne. Ak si zvolíme najstarší vek generácie 100 rokov, potrebujeme ešte počkať 5 rokov, kým sa nám skompletizuje aspoň jedna generácia, a to generácia 1920.

Graf 3. Časovo generačný priestor vypočítateľných pravdepodobností úmrtia v roku 2016



Postupné skompletizovanie generačných pravdepodobností úmrtia je znázornený na grafe 4. Každý rok pribudnú pre každú generáciu, ktorá nie je skompletizovaná, dáta pre ďalší vek.

Graf 4.: Ročná aktualizácia pravdepodobností úmrtia pre generácie

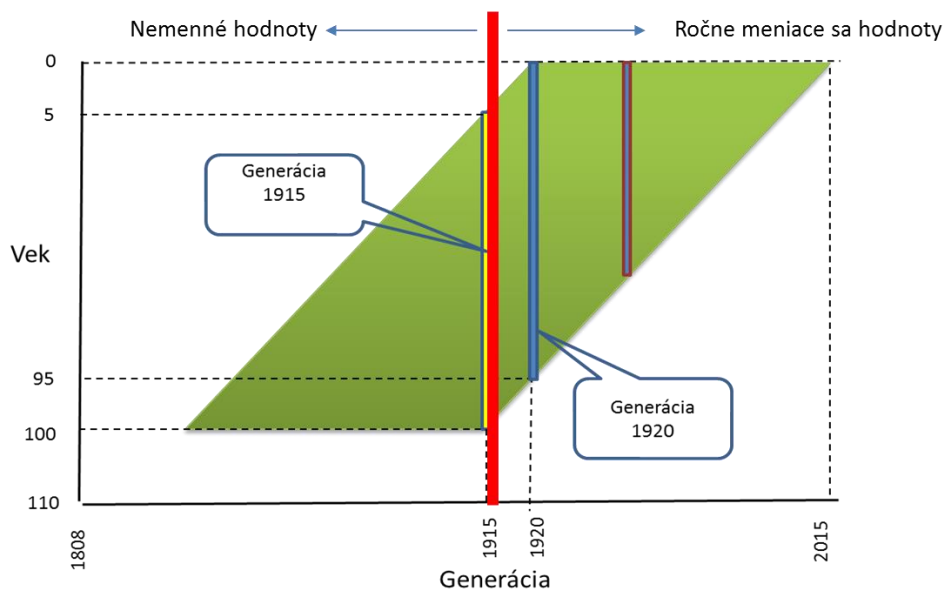


Po výpočte pravdepodobností úmrtí možno pristúpiť na výpočet ďalších funkcií generačných úmrtnostných tabuliek. Je to vytvorenie *počtov tabuľkových žijúcich* pre jednotlivé veku hypotetickej populácie, *tabuľkový počet zomrelých* pre jednotlivé veku, *priemerné počty rokov prežitých tabuľkovými žijúcimi* pre jednotlivé roky, *kumulatívne počty rokov prežitých tabuľkovými žijúcimi* a *priemerné dĺžky života*. Počíta sa to na základe modelu stacionárnej populácie, ktorá by vznikla pri použití skutočných pravdepodobností úmrtia za predpokladu, že sa narodí konštantný počet detí v každom roku, čím sa odstráni vplyv vekovej štruktúry skutočnej skúmanej populácie. Spôsob výpočtu je uvedený v [2]. Z algoritmov vyplýva, že tabuľkový počet žijúcich v jednotlivých vekoch je priamo úmerný s pravdepodobnosťami úmrtia, tým je zrejmé, že sa dajú vypočítať len pre tie veku, kde sú známe pravdepodobnosti.

Z tohto môžeme konštatovať, že **generačné úmrtnostné tabuľky sú neúplné**. Jedna časť, a to generácie pred rokom 1916, sú neúplné pre mladšie veku smerom k veku 0 a aj zostanú neúplné, lebo relevantné dáta sa nedajú zistiť. Generácie 1916 až 1919 sú neúplné pre mladšie veku a aj pre staršie veku. Pre mladšie veku údaje nie je možné získať, ale pre staršie veku do 100 rokov budú postupne dopĺňané. Generácie po roku 1919 sú nekompletné len pre staršie veku a budú postupne skompletizované. O päť rokov vznikne prvá kompletná generačná úmrtnostná tabuľka pre ročník 1920.

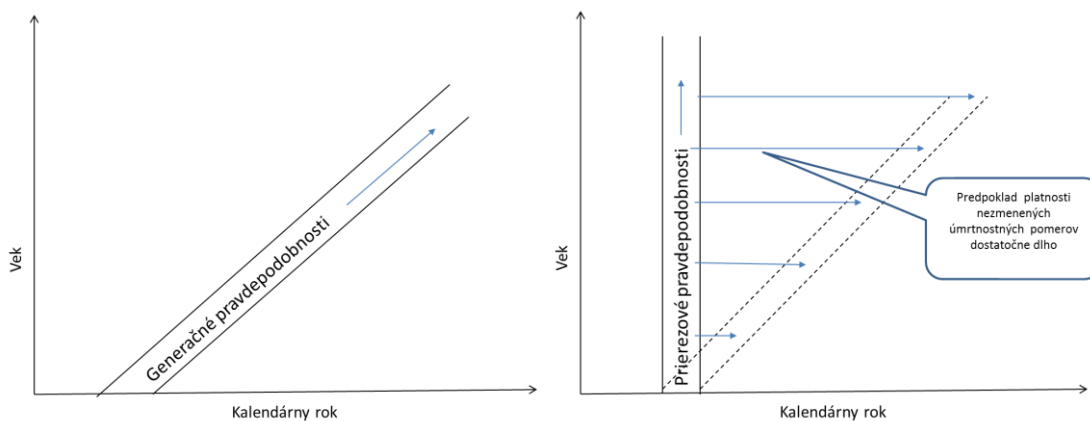
Ďalšia vlastnosť generačných úmrtnostných tabuliek je, že niektoré sú **nemenné** a niektoré sú každý rok **menené**. Nemenné generačné tabuľky sú tie, u ktorých je známa pravdepodobnosť úmrtia v najvyššom veku. V našich podmienkach sú to generácie pred rokom 1916. U týchto generácií môžeme všetky funkcie týchto tabuliek využiť plnohodnotne. Generácia 1916 sa stane nemenná, keď sa doplnia údaje za vek 100 rokov. Je zrejmé, že každý nasledujúci kalendárny rok sa stane nemenná jedna generácia. Treba si uvedomiť, že generácie, ktoré nie sú nemenné neposkytujú plnohodnotné tabuľkové funkcie. Vzťahujú sa len na tú časť populácie, ktorá zomrela do najvyššieho zisteného veku. Napr. v roku 2016 generácia 1960 bola skompletizovaná do veku 55. Priemernú dĺžku života pri narodení, ktorá má hodnotu 52,07 roka musíme interpretovať ako „priemerný vek, ktorého sa dožili osoby generácie 1960, ktorí zomreli do veku 55 rokov“.

Graf 5. Nemenné a meniace sa funkcie generačných úmrtnostných tabuliek



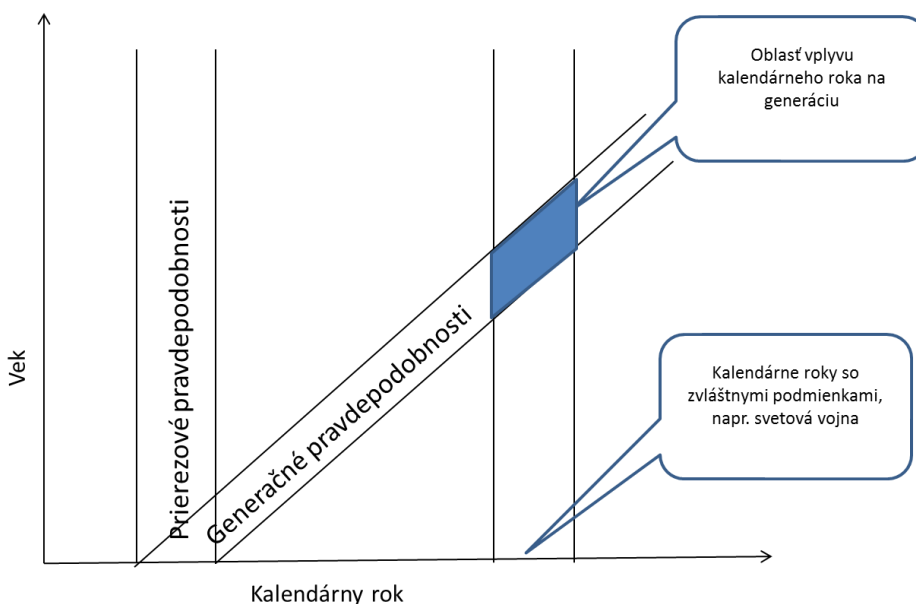
Napriek tomu, že je problém vytvoriť kompletné generačné úmrtnostné tabuľky, ich význam to neznižuje. Dokážu zohľadniť rôzne zdravotné a životné podmienky pre danú generáciu, čo prierezové úmrtnostné tabuľky nedokážu.

Graf 5. Rozdielny výpočet prierezových a generačných pravdepodobností



Ako z grafu 5 vidno, prierezové pravdepodobnosti úmrtia, ktoré budú použité podľa metodiky výpočtu úmrtnostných tabuliek na hypotetickú generáciu o konštantnej veľkosti (napr. 100000 osôb), podliehajú podmienke o nemennosti minimálne počas rokov určujúcich vek osôb na ktoré sa vzťahujú. Čiže, hypotetickú generáciu ovplyvňujú pravdepodobnosti z jedného kalendárneho roka, a to kalendárneho roka, kedy sa hypotetická generácia narodila. Na druhej strane generačné pravdepodobnosti úmrtia sú vypočítané zo skutočných údajov za kalendárne roky, ktoré určujú vek osôb na ktoré sa vzťahujú. Inak povedané na výpočet prierezových pravdepodobností úmrtia nevlývajú údaje získané za kalendárne roky so zvláštnymi podmienkami úmrtí, kým na generačné pravdepodobnosti úmrtia priamo vplývajú.

Graf 6. Vplyv kalendárnych rokov so zvláštnymi podmienkami na výpočet pravdepodobností úmrtia



Aby sa dali využiť hore uvedené vlastnosti generačných úmrtnostných tabuliek, je potrebné vyriešiť úplnosť týchto tabuliek. Ako už bolo písané, doposiaľ nemáme generáciu, ktorá by mala úplnú úmrtnostnú tabuľku, čiže ešte nevymrela úplne, tým nemôžeme vypočítať potrebné ukazovatele pre všetky veku. Je potrebné postupne počkať na dáta z nasledujúcich

kalendárnych rokov, z ktorých údaje tieto tabuľky postupne doplnia. Je toto však časovo náročné.

Existujú postupy, ktoré by umožnili tieto chýbajúce dáta odhadnúť. Základom týchto postupov je odhad chýbajúcich pravdepodobností úmrtia. Pre Slovenské podmienky v súčasnosti sú to pravdepodobnosti pre generácie od 1920 pre veku 95 až 100 rokov, pre generáciu 1921 pre veku 94 až 100 rokov, atď. , až pre generáciu 2015 pre veku 1 až 100 rokov.

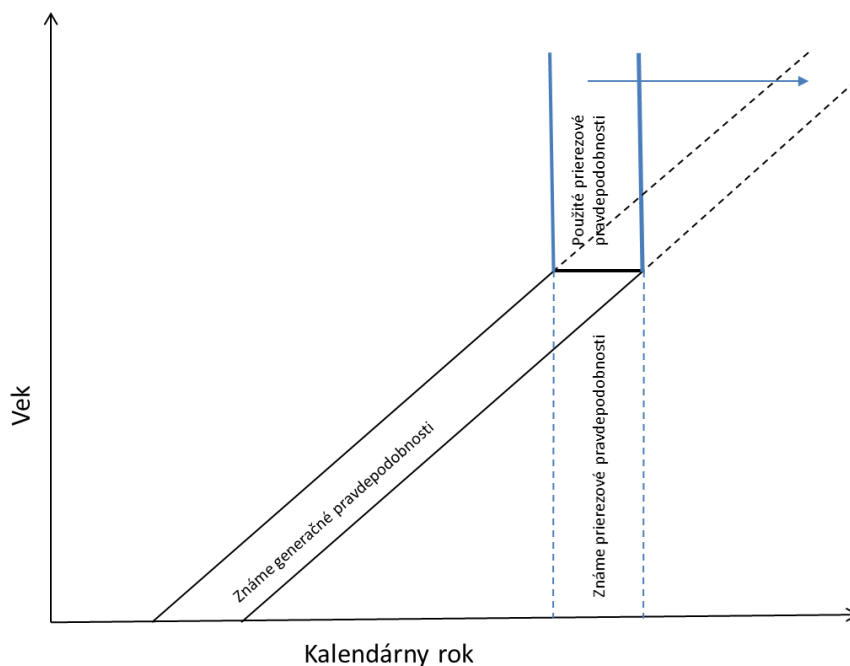
Tieto odhady možno vykonať napríklad jednoduchým matematickým modelom:

$$q_x = f(x)$$

kde f môže byť lineárna, alebo nelineárna funkcia. Funkcia f môže byť napr. jedna z funkcií popísaných [2]. Parametre tejto funkcie možno odhadnúť na základe známych hodnôt pravdepodobností. Nevýhodou takéhoto postupu je, že sa dá použiť len na odhad malého počtu hodnôt pravdepodobností.

Metóda, ktorú použili v [1] sa zakladá na použití pravdepodobností úmrtia počítané v prierezových tabuľkách. Graf 7 znázorňuje nahradenie chýbajúcich generačných pravdepodobností úmrtia s prierezovými pravdepodobnosťami úmrtia.

Graf 7. Metóda odhadu chýbajúcich generačných pravdepodobností s využitím prierezových pravdepodobností



Pri použití tejto metódy je treba poznamenať, že funkcie generačných úmrtnostných tabuliek, ktoré boli vytvorené, sú skreslené lineárne počtom použitých prierezových pravdepodobností. Napr. ak pre generáciu 1920 použijeme prierezové pravdepodobnosti úmrtia pre veku 95 až 100 rokov, dostaneme generačné priemerné dĺžky života menej skreslené, ako keď pre generáciu 2000 použijeme prierezové údaje od veku 16 až po vek 100 rokov. Tento fakt treba pri použití generačných úmrtnostných tabuliek brať do úvahy. Napr. pri hodnotení strednej

dĺžky života pri narodení je zrejmé, že ženy žijú dlhšie muži, tak výber generácie, kde ešte akceptujeme skreslenia bude rozdielny. Jeden z možností nastavenia kritéria akceptovateľnosti skreslenia, je určenie generácie, kde kumulatívna početnosť tabuľkových zomrelých vypočítaných z existujúcich skutočných generačných dát neklesne pod 10% z hodnoty určujúcej koreň tabuľky, v našom prípade pod 10000 zomrelých. Toto kritérium nám hovorí, že pre dáta známe do roku 2016 u mužov môžeme akceptovať odhadnuté generácie od 1920 po 1968 a u žien od 1920 po 1958. Je treba podotknúť, že ak získame dáta pre roky 2017 a vyššie, toto akceptovateľnosť sa bude meniť.

Na záver môžeme zhrnúť **význam generačných úmrtnostných tabuliek** nasledovne. Napriek problémom pri tvorbe a použití generačných úmrtnostných tabuliek treba zdôrazniť ich potrebu. V prvom rade je to možnosť zisťovať vplyv kalendárnych rokov s zvláštnymi podmienkami na úmrtnosť populácie pre jednotlivé veku a na priemernú dĺžku života celej populácie. Možno napr. zistiť vplyv svetových vojen na dĺžku života populácie, vplyv kalendárnych rokov s významnými epidémiami, vplyv kalendárnych rokov s politickými zmenami, atď. Je možné zistiť či prierezové úmrtnostné tabuľky správne odhadujú stredné dĺžky života, a to porovnať prierezové úmrtnostné tabuľky pre určitý kalendárny rok s generačnou úmrtnostnou tabuľkou populácie, ktorá sa narodila v tomto kalendárnom roku. Prierezové stredné dĺžky života sú výsledkom výpočtov za predpokladu platnosti úmrtnostných pomerov, vyjadrených pravdepodobnosťami úmrtí, dostatočne dlho (asi 100 rokov), kým generačné priemerné dĺžky života sú priamym výsledkom výpočtov na skutočných pravdepodobnosti úmrtia. Týmto sa zistí, že podmienka nemennosti úmrtnostných pomerov dostatočne dlho, ako skreslil priemernú dĺžku života populácie.

Literatúra

- [1] *Human Mortality Database*. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). (www.mortality.org alebo www.humanmortality.de)
- [2] Mészáros, J.: Požiadavky modernizácie demografickej štatistiky vo vzťahu k novým metódam graduácie pravdepodobností úmrtia. Metodický materiál. INFOSTAT Bratislava, 2016. (www.infostat.sk/vdc/pdf/TeoriaUT.pdf)
- [3] Mészáros, J.: Výpočet úmrtnostných tabuliek. INFOSTAT Bratislava 2000. (http://www.infostat.sk/vdc/pdf/metodika_ut.pdf)
- [4] Mészáros, J.: Teoretické a metodologické prístupy konštrukcie generačných úmrtnostných tabuliek a ich zhodnotenie v kontexte dátových podmienok Slovenska. Metodický materiál. INFOSTAT Bratislava, 2017