

Beata Jackowska

Uniwersytet Gdański

WPLYW WYBRANYCH METOD WYGŁADZANIA PRAWDOPODOBIENSTW ZGONU NA WYSOKOŚĆ SKŁADEK I REZERW NETTO W UBEZPIECZENIACH NA ŻYCIE

Streszczenie: W ubezpieczeniach na życie konstrukcja tablic trwania życia (tablic wymieralności) ma podstawowe znaczenie dla wyników kalkulacji aktuarialnych. Tablice trwania życia są budowane najczęściej na podstawie danych przekrojowych, a następnie wygładzane w poszukiwaniu nieznanego, ukrytego wzorca wymierania. W artykule porównano na przykładzie empirycznym wybrane metody wygładzania prawdopodobieństw zgonu. Następnie dokonano przykładowych kalkulacji składek i rezerw w przypadku funkcji tablicowych surowych i wygładzonych.

Słowa kluczowe: tablice trwania życia, wygładzanie wskaźników demograficznych, splajny, ubezpieczenia na życie, składki i rezerwy netto.

1. Wstęp

Znajomość rozkładu dalszego trwania życia w populacji osób ubezpieczonych leży u podstaw wszelkich kalkulacji aktuarialnych dotyczących polis ubezpieczeń na życie. W praktyce rozkład dalszego trwania życia wyznacza się na podstawie danych statystycznych, konstruuje tablice trwania życia (tablice wymieralności). Funkcje tablicowe są estymowane najczęściej na podstawie danych przekrojowych, tzn. danych o zgonach i liczbie osób w ustalonym krótkim okresie w badanej populacji (dane te stanowią więc przekrój przez wszystkie żyjące w danym okresie generacje). Przekrojowa tablica wymieralności jest więc krótkotrwałym odzwierciedleniem poziomu umieralności wszystkich ludzi badanej populacji żyjących w danym okresie.

Estymacja funkcji tablicowych dokonywana jest z określoną dokładnością. Im większa jest liczebność danego rocznika wieku, tym większa dokładność. Niestety, zwykle trudno uzyskać zadowalającą dokładność, dlatego konieczne może się stać wygładzanie funkcji tablicowych w poszukiwaniu nieznanego wzorca wymierania.

W artykule na przykładzie empirycznym przedstawiono wpływ wybranych metod wygładzania na poziom bieżącej składki netto oraz rezerwę składki netto, a także zysk (stratę) ze względu na różnicę w śmiertelności.

2. Porównanie wybranych metod wygładzania prawdopodobieństw zgonu

Wygładzanie (wyrównywanie) wskaźników demograficznych polega na zastąpieniu wartości surowych przez wartości bardziej regularne i lepiej opisujące, według arbitralnej opinii badacza, istotę zjawiska. Wybór metody wyrównywania jest dążeniem do kompromisu między osiągnięciem odpowiedniego stopnia wygładzenia i jednocześnie otrzymaniem wartości zbliżonych do wyjściowych.

Wygładzaniu podlega najczęściej jedna z następujących funkcji tablicowych¹: l_x – liczba dożywających wieku x , d_x – liczba osób, które umarły w wieku x , p_x – prawdopodobieństwo, że osoba w wieku x dożyje do wieku $x+1$, q_x – prawdopodobieństwo zgonu w wieku x , m_x – współczynnik zgonów w wieku x .

Polskie tablice trwania życia ogłaszane przez GUS są wygładzonymi tablicami przekrojowymi konstruowanymi na podstawie danych o zgonach i liczbie ludności dla jednego roku kalendarzowego. W celu niniejszej analizy zbudowano dla roku 2007 tablice surowe zgodnie z metodologią stosowaną przez GUS [*Trwanie życia...* 2008, s. 34-35]. Wykorzystano następujące dane:

- 1) liczba osób zmarłych w 2007 r., sklasyfikowanych według płci, roku urodzenia i wieku ukończonego;
- 2) liczba ludności według płci i roczników wieku na koniec 2006 oraz 2007 r.;
- 3) liczba urodzeń według płci w 2007 r.

Spośród całego bogactwa metod do analizy wybrano metodę wygładzania P-splajnami, która została porównana z metodą mechaniczną stosowaną przez GUS do wyrównywania prawdopodobieństw zgonu w polskich tablicach trwania życia.

2.1. Metodologia GUS wygładzania tablic surowych

Prawdopodobieństwa zgonu q_x otrzymane na podstawie danych surowych dla $x = 1, 2, \dots, 84$ są wyrównywane trzykrotnie za pomocą ruchomych średnich ważonych o różnej liczbie wyrazów (5, 7, 9) w zależności od grupy wieku. Dla wieku $x = 0$ prawdopodobieństwo q_0 pozostaje niezmiennione. Jest to metoda mechaniczna, w której liczba uśrednianych wartości, system wag i liczba powtórzeń procedury zostały dobrane na podstawie badań empirycznych ludności Polski. Dla wieku powyżej 85 roku życia stosuje się analityczną metodę wygładzania polegającą na ekstrapolacji funkcji wielomianowo-wykładniczej dla liczby dożywających l_x i na tej podstawie wyznacza się dopiero prawdopodobieństwa zgonu² [*Trwanie życia...* 2008, s. 35-36].

¹ Pozostałe funkcje tablicowe wyznacza się na podstawie wygładzonego wskaźnika demograficznego.

² Więcej na temat mechanicznych i analitycznych metod wygładzania wskaźników demograficznych można znaleźć w [London 1985; Tabeau, van den Berg Jeths, Heathcote 2001; Wieniecki 1976].

2.2. Wygładzanie P-splajnami

Metoda wygładzania funkcjami sklejanymi³ polega na lokalnej interpolacji wskaźników surowych wielomianami sklejanymi w taki sposób, by globalnie zachować pewien stopień gładkości, tzn. różniczkowalność określoną liczbę razy [Eilers, Currie, Durban 2006; Eilers, Marx 1996]. B-splajny są funkcjami bazowymi składającymi się z segmentów gładko połączonych wielomianów (w praktyce zwykle niskiego stopnia drugiego lub trzeciego). Miejsca połączenia segmentów nazywane są węzłami (*knots*). Przy wygładzaniu B-splajnami należy podjąć decyzję o: stopniu wielomianów, liczbie węzłów⁴, położeniu węzłów (zwykle stosuje się węzły równoodległe). Wybór ten ma wpływ na stopień wygładzenia i dopasowania modelu. Liniowa kombinacja B-splajnow daje krzywą wygładzoną. B-splajny tworzą bazę funkcji wykorzystywaną w następującym zagadnieniu regresji (*B-spline regression*):

$$y = B\alpha + e, \quad (1)$$

gdzie: y – wektor $m \times 1$ zmiennej objaśnianej,
 x – wektor $m \times 1$ zmiennej objaśniającej,
 B – macierz $m \times n$ wartości n splajnow wyznaczonych dla argumentu x ,
 α – wektor $n \times 1$ współczynników regresji,
 e – wektor $m \times 1$ reszt.

W tym przypadku zmienną objaśnianą jest wskaźnik demograficzny, natomiast zmienną objaśniającą jest wiek. Współczynniki regresji wyznacza się, minimalizując sumę kwadratów:

$$S = \|y - B\alpha\|^2. \quad (2)$$

P-splajny są liniową kombinacją B-splajnow oraz tzw. funkcji straty (*penalized function*). Funkcja straty oparta jest na różnicach określonego rzędu dla współczynników α sąsiadujących B-splajnow. Rząd różnic jest wybierany w zależności od zastosowań. Współczynniki regresji (*P-spline regression*) wyznacza się, minimalizując sumę kwadratów:

$$S = \|y - B\alpha\|^2 + \lambda \|D_d \alpha\|^2, \quad (3)$$

gdzie: $D_d \alpha$ – wektor $(n-d) \times 1$ różnic rzędu d pomiędzy sąsiednimi współczynnikami regresji $\Delta^d \alpha_i = \Delta^{d-1} \alpha_{i+1} - \Delta^{d-1} \alpha_i$ (dla $i = 1, 2, \dots, n-d$),
 λ – waga funkcji straty.

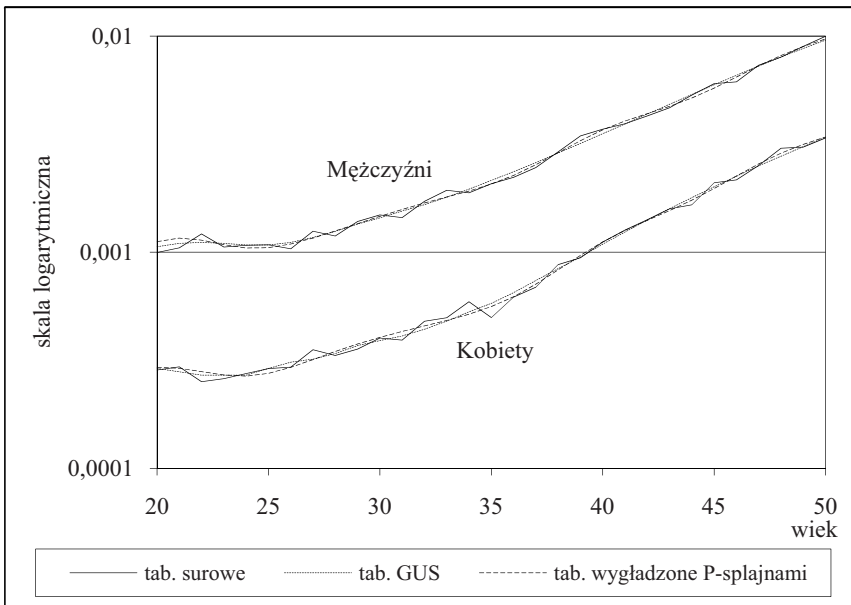
³ Zamiast nazwy funkcje sklepane używa się także określenia funkcje gięte lub splajny (*spline*).

⁴ Im większa liczba węzłów, tym lepsze dopasowanie modelu, lecz krzywa wówczas bardziej „faluje”, odwzorowując nieregularności w danych. Problem algorytmu poszukiwania optymalnej liczby węzłów nie jest rozstrzygnięty i jest przedmiotem dyskusji [Eilers, Currie, Durban 2006; Eilers, Marx 1996].

Zagadnienie regresji w przypadku P-splajnow można rozszerzyć do uogólnionego modelu regresji (GLZ), w którym współczynniki regresji wyznacza się metodą największej wiarygodności, modyfikując funkcję wiarygodności $L(\alpha)$ następująco:

$$L^*(\alpha) = L(\alpha) - \frac{1}{2} \lambda \|D_d \alpha\|^2. \quad (4)$$

W artykule posłużono się metodą wygładzania P-splajnami do wyrównania logarytmu prawdopodobieństw zgonu w wieku $x = 0, 1, \dots, 100$, przyjmując: wielomiany stopnia trzeciego, 25 równoodległych węzłów, różnice rzędu trzeciego dla funkcji straty. Waga λ funkcji straty została wyznaczona za pomocą kryterium krzyżowej poprawności (*cross validation*). Współczynniki regresji wyznaczone metodą największej wiarygodności. Na rys. 1 zostały porównane dla wieku od 20 do 50 lat surowe wartości q_x oraz wyrównane pochodzące z polskich tablic trwania życia [Trwanie życia... 2008] i wygładzone metodą P-splajnow.



Rys. 1. Estymatory surowe i wygładzone prawdopodobieństwa zgonu w wieku od 20 do 50 lat w 2007 r.

Źródło: opracowanie własne.

Oceniając efekt wyrównywania, należy mieć na uwadze, że polepszenie stopnia wygładzenia często odbywa się kosztem pogorszenia dopasowania wielkości wygładzonych do surowych. Dlatego też ocena gładkości powinna być przeprowadzana łącznie z analizą dopasowania [London 1985]. Przy ocenie wygładzenia można się

kierować następującymi kryteriami opartymi na obserwacji różnic określonego rzędu między sąsiednimi wyrazami wyrównywanego ciągu wartości:

- różnice rzędu pierwszego, drugiego itd. powinny być coraz mniejsze,
- znaki różnic nie powinny ulegać częstym zmianom,
- jako miary wygładzenia można przyjąć: sumę wartości bezwzględnych różnic, średnią wartość bezwzględnych różnic, średnią wartość bezwzględnych różnic do średniej wartości szeregu.

Tablice wygładzone opublikowane przez GUS oraz tablice wygładzone P-splajnami spełniają dwa pierwsze kryteria. Miary stopnia wygładzenia (tab. 1) przyjmują zbliżone wartości, przy czym nieco lepsze jest wygładzenie tablic GUS.

Tabela 1. Ocena wygładzenia prawdopodobieństw zgonów na podstawie różnic rzędu d dla osób w wieku od 20 do 60 lat

Miara wygładzenia	Metodologia GUS			Wygładzanie P-splajnami		
	$d = 1$	$d = 2$	$d = 3$	$d = 1$	$d = 2$	$d = 3$
	Mężczyźni					
Suma wartości bezwzględnych różnic	0,02036	0,00159	0,00045	0,02043	0,00205	0,00114
Średnia wartość bezwzględnych różnic do średniej wartości szeregu	8,15%	0,64%	0,18%	8,18%	0,82%	0,46%
	Kobiety					
Suma wartości bezwzględnych różnic	0,00813	0,00064	0,00050	0,06318	0,00879	0,00145
Średnia wartość bezwzględnych różnic do średniej wartości szeregu	9,10%	0,72%	0,56%	9,19%	0,95%	0,60%

Źródło: obliczenia własne.

Jako kryterium oceny dopasowania wygładzonych współczynników demograficznych do współczynników surowych można zaproponować:

- znaki odchyłeń wartości surowych od wygładzonych nie mogą tworzyć zbyt długich serii,
- miary dopasowania: suma kwadratów różnic (lub wartości bezwzględnych różnic) pomiędzy wartościami wygładzonymi a surowymi, odchylenie standardowe reszt, odchylenie standardowe reszt przypadające na średnią wartość szeregu.

Dopasowanie obu rozważanych modeli jest bardzo dobre, przy czym tym razem nieco lepsze jest dopasowanie modelu wygładzonego P-splajnami (tab. 2). Na podstawie miar wygładzenia i dopasowania (tab. 1 i 2) można stwierdzić, że miary absolutne są mniejsze w przypadku szeregu dla kobiet, natomiast miary względne są mniejsze dla mężczyzn⁵.

⁵ Wynika to z tego, że umieralność kobiet jest na znacznie niższym poziomie niż mężczyzn (tzw. nadumieralność mężczyzn).

Tabela 2. Ocena dopasowania wygładzonych prawdopodobieństw zgonów dla osób w wieku od 20 do 60 lat

Miara dopasowania	Mężczyźni		Kobiety	
	metodologia GUS	wygładzanie P-splajnami	metodologia GUS	wygładzanie P-splajnami
Odchylenie standardowe reszt	0,000167	0,000143	0,000116	0,000100
Odchylenie standardowe reszt do średniej wartości szeregu	2,74%	2,35%	5,34%	4,59%

Źródło: obliczenia własne.

Wartość miar wygładzenia i dopasowania zależy nie tylko od metody wygładzania, ale także od gładkości i regularności szeregu surowego. Dlatego też miary te nie niosą informacji o stopniu dopasowania pojedynczego modelu, lecz służą do analizy porównawczej kilku modeli zbudowanych na podstawie tych samych danych.

3. Przykładowe kalkulacje składek i rezerw w przypadku prawdopodobieństw zgonu surowych i wygładzonych

Zaprezentowane powyżej metody zostały zastosowane na dużym zbiorze danych charakteryzujących się regularnością, dlatego też otrzymano bardzo dobrą ocenę wygładzania i dopasowania. Mimo to odchylenia wartości rzeczywistych od modelowych przenoszą się w widoczny sposób na wyniki kalkulacji składek i rezerw w ubezpieczeniach zawieranych na krótki okres. Ponieważ w obu metodach wygładzania spełniony był warunek losowości odchyżeń, więc wygładzanie nie wpływa istotnie na poziom składek i rezerw w ubezpieczeniach długoterminowych. W tab. 3 zaprezentowano względną zmianę jednorazowej składki netto po wygładzeniu prawdopodobieństw zgonu dla przykładowych terminowych ubezpieczeń na wypadek śmierci na okres 3 i 5 lat dla osób w wieku od 20 do 60 lat przy założonej technicznej stopie procentowej 3%⁶.

Natomiast w tab. 4 przedstawiono zmianę wartości w czasie rezerwy składki netto przypadającej na jedną polisę przy zastosowaniu danych surowych i wygładzonych dla przykładowego 10-letniego ubezpieczenia na wypadek śmierci na sumę 1000 jednostek pieniężnych dla mężczyzny i kobiety w wieku 40 lat przy składce płatnej co rok z góry i technicznej stopie procentowej 3%. Rezerwa składki netto (nazywana też rezerwą matematyczną) w momencie t została policzona jako różnica między wartością przyszłych świadczeń zdyskontowanych na moment t oraz wartością przyszłych składek zdyskontowanych na moment t [Bowers i in. 1997].

⁶ Względne odchylenie nie zależy od kwoty ubezpieczenia.

Tabela 3. Względne odchylenie wartości jednorazowej składki netto skalkulowanej dla tablicy wygładzonej od wartości jednorazowej składki netto skalkulowanej dla tablicy surowej przy założeniu technicznej stopy procentowej 3% dla osób w wieku od 20 do 60 lat

Odchylenie względne	Mężczyźni		Kobiety	
	metodologia GUS	wygładzanie P-splajnami	metodologia GUS	wygładzanie P-splajnami
3-letnie ubezpieczenie na wypadek śmierci				
Minimalne	0,084%	0,059%	0,021%	0,077%
Maksymalne	5,244%	5,092%	9,519%	6,965%
Średnie	1,165%	0,804%	1,674%	0,900%
5-letnie ubezpieczenie na wypadek śmierci				
Minimalne	0,039%	0,005%	0,139%	0,044%
Maksymalne	3,523%	3,128%	4,590%	3,079%
Średnie	0,615%	0,303%	1,183%	0,575%

Źródło: obliczenia własne.

Tabela 4. Rezerwa składki netto w momencie t przy założeniu technicznej stopy procentowej 3% w przypadku 10-letniego ubezpieczenia na wypadek śmierci na sumę 1000 dla osoby 40-letniej przy składce płatnej co rok z góry

Moment t	Mężczyźni			Kobiety		
	rezerwa	zmiana względna rezerwy po wygładzeniu		rezerwa	zmiana względna rezerwy po wygładzeniu	
	dane surowe	metodologia GUS	wygładzanie P-splajnami	dane surowe	metodologia GUS	wygładzanie P-splajnami
0	0,0000			0,0000		
1	1,9676	9,94%	2,50%	0,8195	1,25%	-0,42%
2	3,7836	6,22%	-1,14%	1,5132	1,86%	-0,06%
3	5,3207	3,39%	-2,49%	2,0804	0,96%	0,34%
4	6,5237	0,40%	-2,94%	2,4918	-0,02%	1,42%
5	7,0929	0,28%	-0,14%	2,8465	-5,22%	-1,68%
6	6,9907	1,72%	4,17%	2,7729	-2,89%	2,65%
7	6,7696	-4,42%	0,02%	2,6339	-7,09%	-0,67%
8	5,3493	-4,25%	1,31%	2,1289	-9,30%	-2,85%
9	3,2070	-6,78%	-1,48%	1,0981	2,24%	8,34%
10	0,0000			0,0000		

Źródło: obliczenia własne.

W rzeczywistym portfelu polis wystawionych dla x -latków zaobserwowana w k -tym roku ubezpieczenia liczba zmarłych D_k różni się od przyjętej do kalkulacji teoretycznej liczby zgonów $N_k \cdot q_{x+k}$, gdzie N_k jest liczbą polis w portfelu w k -tym

roku. Wskutek odchylenia $N_k \cdot q_{x+k} - D_k$ na koniec roku (w momencie $k + 1$) ryzyku podlega więc kwota netto (*net amount at risk*) $b_{k+1} - {}_{k+1}V$, gdzie b_{k+1} to wysokość świadczenia oraz ${}_{k+1}V$ wysokość rezerwy w momencie $k + 1$. Zysk (stratę)⁷ ze względu na śmiertelność (*mortality profit*) definiuje się następująco [Błaszczyszyn, Rolski 2004]:

$$I_{k+1} = (N_k \cdot q_{x+k} - D_k)(b_{k+1} - {}_{k+1}V). \quad (5)$$

Tabela 5. Zysk (strata) ze względu na śmiertelność na jedną polisę w momencie t (przy założeniach jak w tab. 4)

Moment t	Mężczyźni		Kobiety	
	metodologia GUS	wygładzanie P-splajnami	metodologia GUS	wygładzanie P-splajnami
0	0,0000	0,0000	0,0000	0,0000
1	-0,1707	-0,0281	-0,0239	0,0025
2	-0,0092	0,1145	-0,0251	-0,0036
3	0,0891	0,1086	-0,0036	-0,0088
4	0,1804	0,0749	0,0026	-0,0291
5	0,0340	-0,1663	0,1324	0,0834
6	-0,0680	-0,2795	-0,0870	-0,1237
7	0,4433	0,3201	0,0895	0,0924
8	-0,0636	-0,0471	-0,0118	0,0415
9	0,0002	0,1398	-0,2502	-0,1548
10	-0,2046	-0,0278	0,0165	0,0934

Źródło: obliczenia własne.

Miarę tę można wykorzystać do oceny *a posteriori* efektów wygładzenia, przyjmując rzeczywisty wzorec wymierania oparty na danych surowych oraz teoretyczny model określony przez wygładzone funkcje tablicowe (tab. 5).

4. Wnioski

Wybór sposobu wyrównywania danych jest arbitralny i zależy od zastosowań. Przy wyborze metody wygładzania wskaźników demograficznych wykorzystywanych

⁷ Jeżeli $I_{k+1} > 0$, to mamy do czynienia z zyskiem (w rzeczywistości było mniej zgonów, niż zakładano), natomiast jeżeli $I_{k+1} < 0$, to mamy do czynienia ze stratą (w rzeczywistości było więcej zgonów, niż zakładano).

w kalkulacjach aktuarialnych należy zwrócić uwagę przede wszystkim na następujące fakty:

- Wybór metody wygładzania wpływa na poziom składek oraz rezerw. Wpływ ten jest silniejszy dla krótszego okresu ubezpieczenia oraz dla małych zbiorów danych.
- Metody mechaniczne pozwalają na kontrolę lokalnych prawidłowości poprzez dobór liczby uśrednianych wartości (zwykle nieparzysta liczba), systemu wag, liczby powtórzeń procedury. W efekcie tego na wartość wyrównywanego wskaźnika demograficznego wpływa ustalona liczba sąsiednich wartości surowych.
- Metoda wygładzania B-splajnami i P-splajnami łączy zalety: lokalnego wygładzania (jak metody mechaniczne) oraz gładkiego sklejenia funkcji interpolujących (w taki sposób, by globalnie zachować określony stopień gładkości na połączeniach). Stopień wygładzenia i dopasowania modelu zależy od stopnia wielomianów, liczby i położenia węzłów oraz postaci funkcji straty.
- Przy wyborze metody wygładzania należy mieć na uwadze, że często polepszenie stopnia wygładzenia odbywa się kosztem pogorszenia dopasowania wielkości wygładzonych do surowych. Z tego powodu ocena gładkości powinna być przeprowadzana łącznie z analizą dopasowania.

Literatura

- Błaszczyszyn B., Rolski T., *Podstawy matematyki ubezpieczeń na życie*, WNT, Warszawa 2004.
- Bowers N., Gerber H., Hickman J., Jones D., Nesbitt C., *Actuarial Mathematics*, The Society of Actuaries, Schaumburg, Illinois 1997.
- Eilers P., Currie I., Durban M., *Fast and compact smoothing on large multidimensional grids*, "Computational Statistics & Data Analysis" 2006, no. 50, s. 61-76.
- Eilers P., Marx B., *Flexible smoothing with B-splines and penalties (with discussion)*, "Statistical Science" 1996, vol. 11, no. 2, s. 89-121.
- London D., *Graduation: the Revision of Estimates*, ACTEX, Winsted 1985.
- Tableau E., van den Berg Jeths A., Heathcote C., *Forecasting Mortality in Developed Countries. Insights from a Statistical, Demographic and Epidemiological Perspective*, Kluwer Academic Publishers, New York 2001.
- Trwanie życia w 2007 r.*, Główny Urząd Statystyczny, Warszawa 2008.
- Wieniecki I., *Metody matematyczne w demografii*, PWN, Warszawa 1976.

THE INFLUENCE OF SELECTED GRADUATION METHODS OF PROBABILITY OF DEATH ON NET PREMIUMS AND RESERVES IN LIFE INSURANCE

Summary: One important problem in life insurance is the construction of life tables (mortality tables), on the basis of which insurance values are calculated. Generally, life tables are built on cross-sectional data and then graduated in search of the unknown, underlying mortality pattern. This paper presents the selected graduation methods of probability of death. The selected methods were compared on an empirical example. The research results enable performing the example calculations of premiums and reserves in the case of ungraduated and graduated table functions.