

Renata Gnitecka, Magdalena Sipa

APROKSYMACJA FUNKCJI INTENSYWNOŚCI ZGONÓW WIELOMIANEM TRZECIEGO STOPNIA ORAZ METODĄ GOMPERTZA-MAKEHAMA – PORÓWNANIE

1. Wstęp

Rozkład trwania życia od lat jest przedmiotem badań osób zajmujących się matematyką aktuarialną. Jednak do tej pory nie stworzono uniwersalnego modelu opisującego czas trwania życia człowieka. Firmy ubezpieczeniowe coraz częściej wprowadzają do swojej oferty ubezpieczenia z możliwością wypłaty świadczenia w momencie zajścia zdarzenia ubezpieczeniowego (po przedstawieniu odpowiednich dokumentów), a nie np. na koniec roku (miesiąca), w którym dane zdarzenie miało miejsce. Operacja taka wymaga określenia analitycznej postaci funkcji intensywności wystąpienia zdarzenia ubezpieczeniowego. W artykule wybrano najczęściej ubezpieczane zdarzenie – śmierć ubezpieczonego.

Celem pracy jest aproksymacja funkcji intensywności zgonów na podstawie danych zawartych w polskich tablicach trwania życia [6] za pomocą wielomianu trzeciego stopnia i porównanie otrzymanych w ten sposób wyników z funkcją intensywności uzyskaną za pomocą prawa Gompertza-Makehama. W przypadku wielomianu uzyskuje się funkcję intensywności w postaci uwikłanej. Autorki zaproponowały wykorzystanie metody punktu środkowego do wyprowadzenia jawnej postaci aproksymowanej funkcji. Propozycja została zilustrowana przykładem przeliczonym dla prawdopodobieństwa zgonu dla mężczyzn w Polsce w roku 2004.

2. Rozkłady intensywności zgonów i prawdopodobieństwa zgonów

W tablicach trwania życia publikowanych przez GUS podaje się oszacowane wartości funkcji prawdopodobieństwa zgonu q_x . Funkcja ta określa prawdopodobieństwo warunkowe zgonu w wieku $(x, x + 1)$ pod warunkiem dożycia wieku x .

Wartości funkcji szacowane są metodą największej wiarygodności [1] i liczone na podstawie zaobserwowanej liczby zgonów i liczebności kohorty w kolejnych latach:

$$\hat{q}_x = \frac{d_x}{l_x},$$

gdzie: d_x – liczba zmarłych w przedziale wiekowym $(x, x + 1)$,

l_x – liczba osób dożywających wieku x .

Funkcja, której aproksymację chcemy przeprowadzić, to funkcja intensywności zgonów, oznaczana $\mu(x)$, czyli taka, że wartość $\mu(x)h + o(x)$ jest warunkowym prawdopodobieństwem zgonu w nieskończenie małym przedziale $(x, x + h)$. Wyprowadzenie wzoru na funkcję intensywności zgonów można znaleźć w pracach [1; 3].

Warunkowe prawdopodobieństwo zgonu osoby w przedziale wiekowym $(x, x + 1)$, pod warunkiem dożycia wieku x , jest równe:

$$q_x = 1 - \exp \left\{ - \int_x^{x+1} \mu(s) ds \right\}. \quad (1)$$

Korzystając z zależności (1) i danych dostępnych w tablicach trwania życia, nie można od razu wyznaczyć intensywności zgonów. W literaturze [1; 5] podaje się pewne rozkłady, zwane prawami wymierania, niestety żadne z nich nie opisują dokładnie rozkładu trwania życia. Są to m.in.:

1. Prawo Gompertza:

$$\mu(t) = Bc^t, \quad B, t > 0, c > 1. \quad (2)$$

2. Prawo Gompertza-Makehama:

$$\mu(t) = A + Bc^t, \quad B, t > 0, A \geq -B, c > 1. \quad (3)$$

3. Prawo Weibulla:

$$\mu(t) = Bt^n, \quad B, n, t > 0. \quad (4)$$

Zauważmy, że prawo Gompertza-Makehama jest uogólnieniem prawa Gompertza. W pracy [4] autorki niniejszego artykułu wykazały, że wyniki uzyskane z wykorzystaniem prawa Weibulla są najsłabsze (tzn. wystąpiły największe rozbieżności między zaobserwowanymi wartościami prawdopodobieństwa zgonu q_x a wartościami aproksymowanymi $\hat{q}_x$, które kształtują się zgodnie z prawem Weibulla). Dlatego w niniejszym artykule dokonano aproksymacji $\mu(x)$ za pomocą prawa Gompertza-Makehama.

Korzystając z danych dostępnych w tablicach trwania życia, nie można od razu wyznaczyć funkcji intensywności zgonów. W niniejszej pracy aproksymacja funkcji intensywności $\hat{\mu}(t)$ sprowadza się do wyznaczenia parametrów A , B i c . W tym celu, korzystając z prawa Gompertza-Makehama i zależności (1), wyznaczono prawdopodobieństwo zgonu postaci:

$$\hat{q}_x^{\text{GM}} = 1 - \exp\left\{\frac{Bc^x(1-c)}{\ln c} - A\right\}. \quad (5)$$

Parametry A , B i c można wyznaczyć, minimalizując różnice między wartościami q_x , wziętymi z tablic trwania życia, a wartościami $\hat{q}_x^{\text{GM}}$.

3. Mierniki różnic między wartościami q_x a aproksymowanymi wartościami $\hat{q}_x$

Mierniki różnic M między wartościami q_x a $\hat{q}_x$ mogą być następujące:

- 1) suma kwadratów odległości (tzw. błąd średniokwadratowy)

$$M_1 = \sum_{i=1}^n (q_i - \hat{q}_i)^2, \quad (6)$$

- 2) suma odległości

$$M_2 = \sum_{i=1}^n |q_i - \hat{q}_i|, \quad (7)$$

- 3) maksimum odległości

$$M_3 = \max_i |q_i - \hat{q}_i|, \quad (8)$$

- 4) średni błąd procentowy

$$M_4 = \frac{1}{n} \sum_{i=1}^n \frac{|q_i - \hat{q}_i|}{q_i}. \quad (9)$$

Mniejsze wartości wyżej wymienionych mierników oznaczają mniejsze różnice między obserwowanymi wartościami prawdopodobieństwa zgonu q_x a wartościami aproksymowanymi $\hat{q}_x$.

4. Przykład

Wykorzystując dostępne dane o liczbie zgonów według płci i wieku za rok 2004, podjęto próbę aproksymacji wartości q_x , a następnie funkcji intensywności

zgonów. Rozważono przypadki, kiedy funkcje $\hat{q}_x$ przybierają postać (5) lub następującą postać wielomianu trzeciego stopnia:

$$\hat{q}_x = D + Cx + Bx^2 + Ax^3. \quad (10)$$

W [4] pokazano, że wyznaczając postać funkcji $\hat{q}_x$ dla wszystkich lub prawie wszystkich danych, otrzymano duże błędy procentowe M_4 i proponowano wyznaczać postać funkcji $\hat{q}_x$ w trzech okresach. Dlatego w niniejszej pracy również wyznaczono oddzielne funkcje $\hat{q}_x$ dla trzech następujących segmentów (przyjętych już w pracy [4]):

- okres 1 (segment 1), obejmujący wiek: 4-19 lat,
- okres 2 (segment 2), obejmujący wiek: 20-50 lat,
- okres 3 (segment 3), obejmujący wiek: 51-100 lat.

Dla każdego segmentu osobno i dla ustalonej postaci funkcji $\hat{q}_x$ wielokrotnie wyznaczano parametry i ostatecznie wybierano ten zestaw parametrów, dla których miernik M_4 przyjął wartość najmniejszą. W przypadku zarówno funkcji Gompertza-Makehama, jak i wielomianu do wyznaczenia parametrów stosowano metody numeryczne. Różnica polegała na tym, że w przypadku wielomianu „wstępne” wartości parametrów ustalono, wykorzystując metodę najmniejszych kwadratów (minimalizując wartość M_1). W każdym przypadku po wyznaczeniu parametrów funkcji $\hat{q}_x$ obliczono dodatkowo wartości wszystkich pozostałych błędów M .

Parametry funkcji $\hat{q}_x$ oraz wartości M_4 i pozostałych mierników M podano w tab. 1 i 2.

Dla wszystkich wyznaczonych segmentów stwierdzono, że aproksymacja wielomianem jest „lepsza” w sensie mniejszego błędu M_4 . Wyraźnie mniejszy błąd M_4 otrzymano dla segmentu pierwszego i drugiego, co sugeruje wstępnie, aby w tych okresach do aproksymacji wykorzystywać wielomian, a nie funkcję Gompertza-Makehama. Jeśli $\hat{q}_x$ jest wielomianem trzeciego stopnia, to nie można podać dokładnej jawnej postaci funkcji intensywności. Dlatego funkcję intensywności zapisano w postaci uwikłanej jako:

$$1 - \exp \left\{ - \int_x^{x+1} \mu(t) dt \right\} = \hat{q}_x = \quad (11)$$

$$= \begin{cases} 3,93 \cdot 10^{-5} + 7,75 \cdot 10^{-5}t - 1,19 \cdot 10^{-5}t^2 + 5,43 \cdot 10^{-7}t^3, & \text{dla } t \in [4, 20), \\ -0,0034787135 + 0,000565752t - 2,42153 \cdot 10^{-5}t^2 + \\ + 3,6177 \cdot 10^{-7}t^3, & \text{dla } t \in [20, 51), \\ -0,563 + 0,0296t - 0,000522t^2 + 3,18 \cdot 10^{-6}t^3, & \text{dla } t \in [51, 100]. \end{cases}$$

Tabela 1. Funkcja Gompertza-Makehama w poszczególnych segmentach (wartości parametrów i błędy M)

Lata	A	B	C	M_1	M_2	M_3	M_4 [%]
4-19	0,000157798	$5,7978 \cdot 10^{-7}$	1,453622457	$2,04222 \cdot 10^{-8}$	0,000388884	0,000116	8,50
20-50	0,000723424	$1,88052 \cdot 10^{-5}$	1,131164414	$7,97966 \cdot 10^{-7}$	0,002591158	0,000709	2,71
51-100	0	$1,68 \cdot 10^{-4}$	1,08	$6,21791 \cdot 10^{-5}$	0,043522269	0,002664	1,95

Źródło: obliczenia własne.

Tabela 2. Wielomian trzeciego stopnia w poszczególnych segmentach (wartości parametrów i błędy M)

Lata	D	C	B	A	M_1	M_2	M_3	M_4 [%]
4-19	$3,93 \cdot 10^{-5}$	$7,75 \cdot 10^{-5}$	$-1,19 \cdot 10^{-5}$	$5,43 \cdot 10^{-7}$	$2,19199 \cdot 10^{-9}$	$8,95 \cdot 10^{-5}$	$4,439 \cdot 10^{-5}$	1,70
20-50	-0,003478135	0,000565752	$-2,42153 \cdot 10^{-5}$	$3,6177 \cdot 10^{-7}$	$9,0538 \cdot 10^{-8}$	0,001278	0,0001353	1,58
51-100	-0,563	0,0296	$-5,22 \cdot 10^{-4}$	$3,18 \cdot 10^{-6}$	0,000498999	0,089649	0,0117728	1,57

Źródło: obliczenia własne.

5. Wyznaczenie jawnej postaci funkcji intensywności

Z równania (11) można wyznaczyć przybliżoną postać funkcji intensywności zgonów $\mu(x)$, stosując metodę punktu środkowego.

Opis metody punktu środkowego można znaleźć np. w [2]. Stosując ją do wzoru (1), otrzymujemy:

$$-\ln(1 - q_x) = \int_x^{x+1} \mu(s) ds = \int_0^1 \mu(x + s) ds \approx \mu(x + 0,5) \quad (12)$$

lub

$$-\ln(1 - q_{x+0,5}) = \int_{x+0,5}^{x+1,5} \mu(s) ds = \int_0^1 \mu(x + 0,5 + s) ds \approx \mu(x + 1). \quad (13)$$

Analogicznie do wzorów (12) i (13) można pokazać, że

$$\mu(x) \approx -\ln(1 - q_{x-0,5}). \quad (14)$$

Po dokonaniu aproksymacji q_x wielomianem i wyznaczeniu paametrów: D , C , B , A (tab. 2) można, wykorzystując (14), zapisać $\mu(x)$ jako:

$$\mu(x) \approx -\ln(1 - D - C(x - 0,5) - B(x - 0,5)^2 - A(x - 0,5)^3). \quad (15)$$

Wartości funkcji intensywności dane wzorem (15) są wartościami przybliżonymi z dwóch powodów:

- metoda punktu środkowego pozwala wyznaczyć $\mu(x)$ na podstawie q_x z pewnym przybliżeniem,
- wartości $q_{x-0,5}$ są wartościami aproksymowanymi, tj. liczonymi przy założeniu, że parametry D , C , B , A wyznaczono, minimalizując różnicę między q_x a $\hat{q}_x$ (minimalizując M_4).

Aby utwierdzić się w wyborze aproksymacji wielomianem, należy jeszcze zbadać, na ile dobre jest przybliżenie $\mu(x)$ dane wzorem (15) dla wyróżnionych segmentów.

W tym celu, całkując¹ $\mu(x)$ daną wzorem (15), wyznaczono warunkowe prawdopodobieństwa zgonu w przedziałach $(x, x + 1)$, które zostały oznaczone dalej jako $\hat{q}_x^*$. Następnie porównano różnice między q_x , $\hat{q}_x^*$ oraz $\hat{q}_x$ licząc błędy M_4 . Wyniki przedstawiono w tab. 3.

Z zebranych w tab. 3 danych wynika, że przybliżenie $\mu(x)$ określone wzorem (15) jest dla badanych segmentów wystarczająco dobre w tym sensie, że $\hat{q}_x^*$ różni

¹ Wartość całki liczono numerycznie, stosując metodę prostokątów.

Tabela 3. Różnice między wartościami q_x , $\hat{q}_x^*$ oraz $\hat{q}_x$ w poszczególnych segmentach

Wyszczególnienie	Segment 1	Segment 2	Segment 3
M_4 [%] między $\hat{q}_x^*$ i $\hat{q}_x$	0,21	0,04	0,02
M_4 [%] między q_x i $\hat{q}_x^*$	1,71	1,58	1,57
M_4 [%] między q_x i $\hat{q}_x$	1,70	1,58	1,57

Źródło: obliczenia własne.

się od q_x praktycznie w takim samym stopniu co $\hat{q}_x$ od q_x . Dodatkowo widać, że błędy między $\hat{q}_x^*$ i $\hat{q}_x$ są w poszczególnych segmentach bardzo małe. Ponadto można zauważyć, że błędy M_4 między q_x i $\hat{q}_x^*$ są wyraźnie mniejsze od błędów M_4 między q_x a prawdopodobieństwem $\hat{q}_x^{\text{GM}}$ obliczonym na podstawie wzoru (5). Przemawia to ostatecznie za zastosowaniem wielomianu trzeciego stopnia do aproksymacji $\mu(x)$.

Alternatywnie, wykorzystując metodę punktu środkowego, można wyznaczyć funkcję intensywności zgonów, posługując się prawdopodobieństwem ${}_2q_x$, czyli prawdopodobieństwem zgonu osoby w przedziale wiekowym $(x, x+2)$ pod warunkiem dożycia wieku x . Prawdopodobieństwo ${}_2q_x$ liczymy z danych zawartych w tablicach trwania życia według wzoru:

$${}_2q_x = \frac{d_x + d_{x+1}}{l_x}. \quad (16)$$

Następnie analogicznie do wzoru (1) mamy:

$${}_2q_x = 1 - \exp \left\{ \int_x^{x+2} \mu(s) ds \right\}. \quad (17)$$

Przekształcając wzór (17) i wykorzystując metodę punktu środkowego, otrzymano funkcję intensywności zgonów postaci:

$$-\ln(1 - {}_2q_x) = \int_x^{x+2} \mu(s) ds = \int_0^2 \mu(x+s) ds \approx 2\mu(x+1). \quad (18)$$

Jest to równoważne z:

$$\mu(x) \approx -0,5 \ln(1 - {}_2q_x). \quad (19)$$

Po dokonaniu aproksymacji ${}_2q_x$ wielomianem trzeciego stopnia i wyznaczeniu parametrów wielomianu D_2, C_2, B_2, A_2 (tab. 7) funkcja $\mu(x)$ przyjmuje następującą postać:

$$\mu(x) \approx -0,5 \ln(1 - D_2 + C_2(x-1) + B_2(x-1)^2 - A_2(x-1)^3). \quad (20)$$

Porównując wyniki zawarte w tab. 2 i 4, można zauważyć, że dla segmentów 2 i 3 istnieje lepsze dopasowanie wielomianu do prawdopodobieństw ${}_2q_x$ niż do q_x (mniejsze wartości błędów w tabeli 4).

Tabela 4. Wielomian trzeciego stopnia wyznaczany na podstawie ${}_2q_x$ (wartości parametrów i błędy M)

Lata	D_2	C_2	B_2	A_2	M_1	M_2	M_3	M_4 [%]
4-19	$3,63 \cdot 10^{-4}$	$5,03 \cdot 10^{-5}$	$-1,26 \cdot 10^{-5}$	$7,56 \cdot 10^{-7}$	$1,15 \cdot 10^{-8}$	0,0003016	$6,58 \cdot 10^{-5}$	2,64
20-50	$-4,96 \cdot 10^{-3}$	$9,32 \cdot 10^{-4}$	$-4,22 \cdot 10^{-5}$	$6,70 \cdot 10^{-7}$	$4,66 \cdot 10^{-7}$	0,0025159	0,000328	1,38
51-99	-0,627	$3,49 \cdot 10^{-2}$	$-6,56 \cdot 10^{-4}$	$4,32 \cdot 10^{-6}$	0,000323	0,0912515	0,005321	1,11

Źródło: obliczenia własne.

Następnie sprawdzono dokładność przybliżenia $\mu(x)$ wzorem (20). W tym celu, całkując $\mu(x)$ daną wzorem (20), wyznaczono warunkowe prawdopodobieństwa zgonu w przedziałach $(x, x+2)$, które zostały oznaczone dalej jako ${}_2\hat{q}_x^*$. Następnie porównano różnice między ${}_2q_x$, ${}_2\hat{q}_x^*$ oraz ${}_2\hat{q}_x$, licząc błędy M_4 . Wyniki przedstawiono w tab. 5.

Tabela 5. Różnice między wartościami ${}_2q_x$, ${}_2\hat{q}_x^*$ oraz ${}_2\hat{q}_x$ w poszczególnych segmentach

Wyszczególnienie	Segment 1	Segment 2	Segment 3
M_4 [%] między ${}_2\hat{q}_x^*$ i ${}_2\hat{q}_x$	0,66	0,14	0,09
M_4 [%] między ${}_2q_x$ i ${}_2\hat{q}_x^*$	3,01	1,39	1,10
M_4 [%] między ${}_2q_x$ i ${}_2\hat{q}_x$	2,64	1,38	1,11

Źródło: obliczenia własne.

Dodatkowo dla każdego segmentu osobno obliczono błąd M_4 między q_x a wartością warunkowego prawdopodobieństwa, oznaczonego $({}_2)\hat{q}_x^*$, określanego w przedziale $(x, x+1)$, ale wyznaczanego po scałkowaniu $\mu(x)$ otrzymanej na podstawie wzoru (20). Wartości tych błędów dla każdego segmentu podano w tab. 6.

Porównując wyniki prezentowane w tab. 3 i 6, można zauważyć, że dla segmentu pierwszego błąd M_4 jest mniejszy, kiedy $\mu(x)$ jest obliczane ze wzoru (15),

Tabela 6. Różnice między wartościami q_x i $({}_2)\hat{q}_x^*$ w poszczególnych segmentach

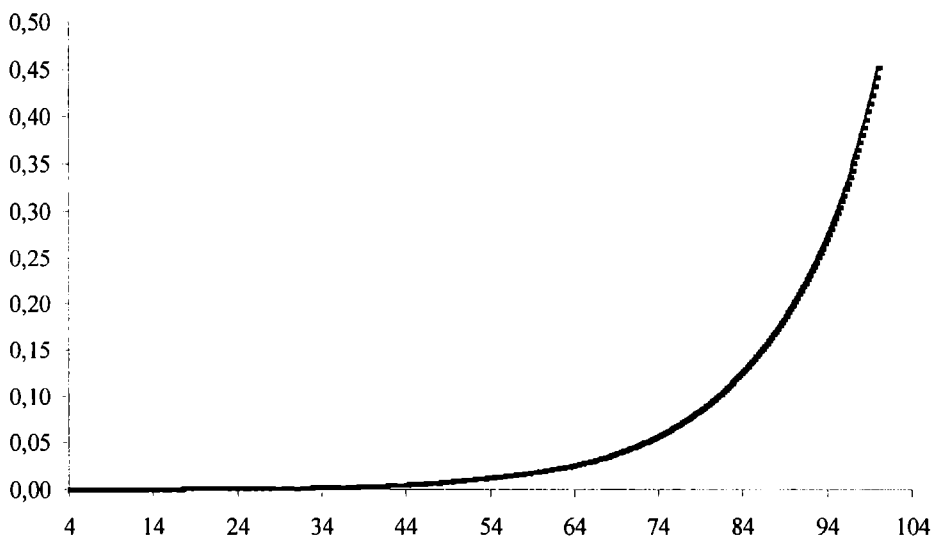
Wyszczególnienie	Segment 1	Segment 2	Segment 3
M_4 [%] między q_x i $({}_2)\hat{q}_x^*$	3,85	1,57	1,19

Źródło: obliczenia własne.

a dla segmentów drugiego i trzeciego błędy M_4 są mniejsze, jeśli $\mu(x)$ jest wyznaczone na podstawie (20). Przyjmując w każdym segmencie najmniejsze wartości błędów M_4 , można więc zaproponować następującą funkcję intensywności zgonów:

$$\mu(x) = \begin{cases} -\ln\left(1 - 0,0000393 - 0,0000775(x-0,5) + 0,0000119(x-0,5)^2 - \right. \\ \quad \left. - 5,43 \cdot 10^{-7}(x-0,5)^3\right), & \text{dla } x \in [2, 20), \\ -0,5\ln\left(1 + 0,00496 - 0,000932(x-1) + 4,22 \cdot 10^{-5}(x-1)^2 - \right. \\ \quad \left. - 6,70 \cdot 10^{-7}(x-1)^3\right), & \text{dla } x \in [20, 51), \\ -0,5\ln\left(1 + 0,627 - 0,0349(x-1) + 0,000656(x-1)^2 - \right. \\ \quad \left. - 4,32 \cdot 10^{-6}(x-1)^3\right), & \text{dla } x \in [51, 100). \end{cases}$$

Funkcję tę przedstawiono na rys. 1.



Rys. 1. Funkcja intensywności dana wzorem (21)

Źródło: opracowanie własne.

6. Wnioski

Obliczenia dokonane na podstawie przykładowych danych o zgonach mężczyzn w roku 2004 w Polsce prowadzą do następujących spostrzeżeń:

1. Aproksymacja z użyciem funkcji Gompertza-Makehama zapewnia stosunkowo niewielkie błędy M_4 w drugim segmencie (lata 20-50) i trzecim segmencie

(lata 51-100), podczas gdy aproksymacja wielomianem charakteryzuje się niewielkimi błędami we wszystkich trzech segmentach.

2. Funkcja intensywności w przypadku wielomianu ma postać uwikłaną, jednak używając metody punktu środkowego, można podać przybliżoną postać funkcji intensywności.
3. Przybliżając $\mu(x)$, można posługiwać się wartościami prawdopodobieństwa dla przedziałów zarówno $(x, x + 1)$, jak i $(x, x + 2)$.
4. Przybliżenie $\mu(x)$ dane wzorem (21) jest dość „dobre”, ze względu na niewielki błąd między wartościami prawdopodobieństwa otrzymanymi na podstawie aproksymowanej $\mu(x)$ a wartościami empirycznymi.

Literatura

- [1] Bolesławski L., *Budowa tablic trwania życia*, GUS, Warszawa 1973.
- [2] Bowers N.L., Gerber H.U., Hickman J.C., Jones D.A., Nesbit C.J., *Actuarial Mathematics*, The Society of Actuaries, Itasca, Illinois 1986.
- [3] *Demografia. Metody analizy i prognozowania*, red. M. Cieślak, PWN, Warszawa 1992.
- [4] Gnitecka R., Sipa M., *Aproksymacja funkcji intensywności zgonów*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1096, *Ekonometria* 15, AE, Wrocław 2005.
- [5] Ostasiewicz S., *Składki w wybranych typach ubezpieczeń życiowych*, AE, Wrocław 2000.
- [6] www.stat.gov.pl.

THE APPROXIMATION OF THE FORCE OF MORTALITY WITH 3RD ORDER POLYNOMIAL AND GOMPERTZ-MAKEHAM FORMULA. A COMPARISON

Summary

Population mortality analysis is being carried out on the basis of the life tables. These tables include the discreet functions (e.g. death probability function) but their continuous equivalents can be approximated (e.g. the force of mortality).

In the paper the approximation of the force of mortality on the basis of the data taken from the Polish life tables for men in 2004 was carried out. To do that the authors used analytical distribution of the future lifetime, Gompertz-Makeham formula and 3rd order polynomial. The authors suggested also using midpoint formula to determine the explicit form of intensity function.

Dr Renata Gnitecka jest adiunktem w Zakładzie Metod Ilościowych Państwowej Wyższej Szkoły Zawodowej w Legnicy.

Mgr inż. Magdalena Sipa jest doktorantką w Katedrze Statystyki i Cybernetyki Ekonomicznej Akademii Ekonomicznej we Wrocławiu.