

Stanisława OSTASIEWICZ\*

## APROKSYMACJA CZASU TRWANIA ŻYCIA W POPULACJACH NIEJEDNORODNYCH

*W demografii jak też naukach aktuarialnych od dawna trwa dyskusja na temat tego, czy istnieje uniwersalne prawo opisujące proces wymierania populacji. Od wielu lat czynione są próby znalezienia takiego prawa. Mimo że proponowano wiele funkcji żadna z nich nie opisuje wystarczająco dokładnie rozkładu empirycznego. Trudność tkwi w tym, że populacja ludzka jest niejednorodna i do opisu umieralności trzeba stosować mieszankę różnych rozkładów.*

*W pracy przedstawiona jest próba opisanie długości życia populacji mężczyzn polskich w roku 2009 za pomocą mieszanki rozkładów Gompertza, Weibula, gamma i lognormalnego. Parametry rozkładów oraz wagi mieszanki wyznaczone będą metodą minimum chi-kwadrat.*

**Słowa kluczowe:** model trwanie życia, śmiertelność, mężczyźni, mieszanka rozkładów, niejednorodność

### WSTĘP

Podstawową wielkością w analizie przeżycia jest czas, jaki upływa do momentu śmierci jednostki. Wielkość ta jest nieujemną zmienną losową, którą zwykle oznacza się symbolem  $T_x$ , gdzie  $x$  jest wiekiem osoby i charakteryzuje się ją za pomocą dystrybuanty:

$$F_x(t) = P(T_0 \leq x+t | T_0 > x) \equiv F(t) = P(T_x \leq t) \quad (1)$$

Wartość dystrybuanty  $F_x(t)$  w punkcie  $t_0$  oznacza prawdopodobieństwo, że osoba w wieku  $x$  przeżyje nie więcej niż  $t_0$  lat. Inną charakterystyką zmiennej losowej  $T_x$  jest funkcja gęstości, która z dystrybuantą związana jest następująco:

$$f_x(t) = F'_x(t) \quad (2)$$

---

\* dr hab. Stanisława OSTASIEWICZ, prof. nadzw. WSOWL – Wydział Zarządzania Wyższej Szkoły Oficerskiej Wojsk Lądowych

gdzie:

$F'_x(t)$  oznacza pochodną funkcji  $F_x(t)$ .

Do charakteryzowania czasu życia wykorzystuje się również funkcję przeżycia  $S_x(t)$ , określoną w następujący sposób:

$$S_x(t) = P(T_0 > x + t | T_0 > x) = 1 - F_x(t) \quad (3)$$

Podstawowym narzędziem analizy czasu przeżycia jest funkcja intensywności, która wyraża się następującym wzorem:

$$\mu(t) \equiv \frac{f_x(t)}{S_x(t)} = -\frac{d}{dt} \ln S_x(t) \quad (4)$$

Funkcja intensywności w sposób jednoznaczny wyznacza funkcję gęstości i dystrybuantę zmiennej losowej  $T_x$ . Wzór (4) można też zapisać następująco:

$$S_x(t) = \exp\left(\int_0^t \mu(s) ds\right) \quad (5)$$

Funkcja intensywności może być interpretowana jako „łatwość” zgonu w chwili  $t$  (na jednostkę czasu) pod warunkiem dożycia do chwili  $t$ , przyjmuje ona wartości z przedziału  $(0, \infty)$ . W prezentowanej pracy funkcja intensywności nie będzie wykorzystywana.

## 1. TEORETYCZNE MODELE CZASU PRZEŻYCIA

Od dawna czynione są próby opisanie długości czasu życia za pomocą znanego rozkładu teoretycznego. Prezentowany artykuł należy do tego typu prac. Jego celem jest opisanie trwania życia populacji mężczyzn polskich w roku 2009.

W rozdziale tym przedstawione są rozkłady teoretyczne wykorzystywane przy realizacji zadania.

Pierwszym parametrycznym modelem życia wykorzystywanym do opisu populacji ludzkich był model zaproponowany w roku 1725 przez de Moivre'a [1]. Czas przeżycia  $T_x$  w tym modelu opisywany był za pomocą rozkładu jednostajnego [10]. W chwili obecnej model ten nie jest stosowany, ma on znaczenie wyłącznie historyczne. Znacznie częściej do modelowania czasu przeżycia wykorzystywany jest model zaproponowany w 1825 roku przez B. Gompertza [2, 4, 7, 8].

Funkcja gęstości i dystrybuanta rozkładu Gompertza są następujące:

$$f(t) = \exp\left\{-\frac{b}{\gamma} \exp((\gamma) - 1) \exp(\gamma)\right\} \quad (6)$$

gdzie:

$$b > 0, \gamma > 0$$

$$F(t) = 1 - \exp\left\{-\frac{\beta}{\gamma} \exp((\gamma) - 1)\right\} \quad (7)$$

Model ten najczęściej zapisywany jest za pomocą funkcji intensywności, która ma postać:

$$\mu(t) = \beta\gamma^t, \quad t \geq 0, \quad \gamma > 0, \quad \beta > 0$$

Taka postać modelu uzasadniana była przez Gompertza w następujący sposób [1]:

Śmiertelność (intensywność zgonów) jest wynikiem działania dwóch rodzajów przyczyn. Pierwszą przyczynę stanowią choroby, które działają tak samo na ludzi młodych, jak i starszych. Tak więc liczba zgonów z tych przyczyn nie zależy od wieku i jest proporcjonalna do liczebności populacji. Drugą przyczyną zgonów to spadek zdolności człowieka do przeciwstawiania się śmierci. Spadek odporności organizmu na śmierć (witalność) jest proporcjonalny do wieku.

Model Gompertza dość dobrze opisuje rzeczywistą intensywność zgonów w grupie wiekowej powyżej 80 lat, jednak dla ludności młodszej zgodność ta jest znacznie gorsza.

Powstało więc przypuszczenie, że oprócz dwóch wymienionych istnieją jeszcze inne przyczyny, które nie zostały uwzględnione przy konstrukcji modelu i które powodują, że funkcja Gompertza nie opisuje dobrze intensywności empirycznej.

Model ten w 1867 r. został zmodyfikowany przez Makehama, który uważał, że intensywność zgonów w każdym wieku jest częściowo niezależna od wieku. Zależy natomiast od pewnych czynników, które oddziałują na intensywność w sposób addytywny niezależnie od wieku. Czynniki te są zagregowane i występują w modelu w postaci stałej, która dodawana jest do funkcji Gompertza [1].

Otrzymany model nazywany jest prawem Gompertza - Makehama i ma postać: dystrybuanta:

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp\left(-ax - \frac{b}{\gamma}(e^{\gamma x} - 1)\right) & x > 0 \end{cases} \quad (8)$$

gdzie:

$$a \geq 0, b > 0, \quad \gamma > 0$$

Rozkład Gompertza jest szczególnym przypadkiem rozkładu Gompertza – Makehama. Jeżeli przyjmiemy  $a = 0$ , otrzymamy rozkład Gompertza.

Gęstość zmiennej losowej ma postać:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ (a + be^{\gamma x}) \exp\left(-ax - \frac{b}{\gamma}(e^{\gamma x} - 1)\right) & x > 0 \end{cases} \quad (9)$$

Zarówno model Gompertza, jak i Gompertza – Makehama wykorzystywane są do dziś. Oprócz tych omówionych modeli stosowanych do opisu przeżycia populacji ludzkiej wykorzystywane też są inne rozkłady teoretyczne, które nawet nazywane są prawami życia, ale żaden z nich nie opisuje dobrze rzeczywistych procesów umieralności.

W dalszej części pracy wykorzystywane będą jeszcze następujące rozkłady: rozkład Weibula, rozkład gamma i rozkład lognormalny.

Dystrybuanty i funkcje gęstości tych rozkładów mają następującą postać:

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right) & x > 0 \end{cases} \quad (10)$$

Funkcja gęstości rozkładu Weibula:

$$f(x) = \begin{cases} 0, & x \leq 0 \\ \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^\alpha\right), & x > 0 \end{cases} \quad (11)$$

Dystrybuanta rozkładu gamma:

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{\Gamma(\alpha, x/\beta)}{\Gamma(\alpha)} & x > 0 \end{cases} \quad (12)$$

gdzie:

$$\Gamma(\alpha, x) = \int_0^x t^{\alpha-1} \exp(-t) dt$$

oraz  $\beta > 0, \alpha > 0$

Funkcja gęstości rozkładu gamma określona jest wzorem:

$$f(x) = \begin{cases} \frac{c^{-a} x^{a-1} \exp(-\frac{x}{c})}{\Gamma(a)}, & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (13)$$

Dystrybuanta rozkładu lognormalnego:

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \Phi\left(\frac{\ln x - m}{\sigma}\right), & x > 0 \end{cases} \quad (14)$$

gdzie:

$$\Phi(t) = \int_0^t \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

Funkcja gęstości rozkładu lognormalnego ma postać:

$$f(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln x - m}{\sigma}\right)^2\right), & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (15)$$

## 2. MODEL EMPIRYCZNY TRWANIA ŻYCIA

Mimo że od dawna czynione są próby skonstruowania funkcji, która mogłaby być traktowana jako prawo opisujące czas życia, to jak do tej pory kończyły się one niepowodzeniem.

Opracowano więc empiryczny model trwania życia, który nazwany został tablicami wymieralności. Później nazwa ta została zmieniona na Tablice Trwania Życia (TTŻ) [12]. W Polsce model ten konstruowany jest ostatnio co roku, osobno dla mężczyzn, osobno dla kobiet. Tablice trwania życia mężczyzn w Polsce w roku 2009 zamieszczone są w pracy jako załącznik 1.

Model ten bazuje na następujących funkcjach biometrycznych [1, 5]:

- liczba dożywających wieku  $x$ ;
- liczba zmarłych w wieku  $x$ ;
- prawdopodobieństwo zgonu w wieku  $x$ .

Wiek  $x$  podawany jest w sposób dyskretny w punktach  $x = 0, 1, 2, \dots, \omega$ .

$\omega$  – jest to najstarsza grupa wieku i w polskich TTŻ jest ona równa 100 lat.

$l_0$  – jest to liczba dożywających wieku 0 lat, czyli początkowa liczebność populacji, w polskich TTŻ wynosi ona 100 000 osób. W miarę upływu czasu następuje wymieranie populacji, w wieku 100 lat liczebność ta jest bardzo mała i tablice nie są kontynuowane.

Najważniejszą funkcją biometryczną są współczynniki zgonów, które wyznaczone są na podstawie współczynników zgonów obserwowanych w okresie dla którego budowane są tablice trwania życia. Na podstawie współczynników zgonów można odtworzyć wszystkie funkcje biometryczne.

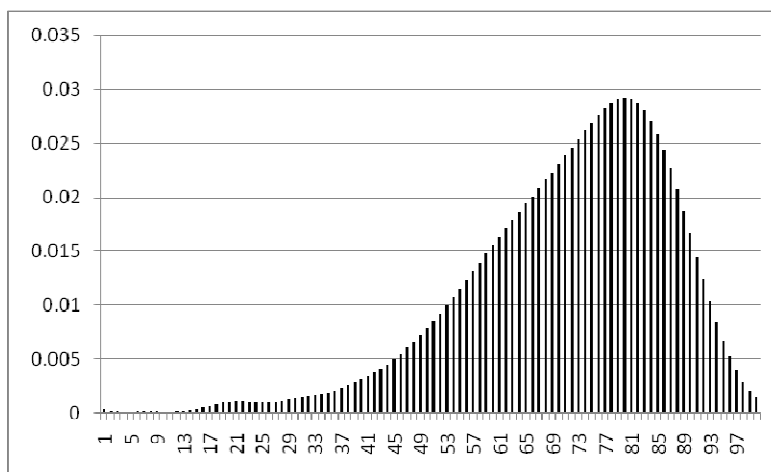
Istnieje związek między empirycznym modelem trwania życia i modelem teoretycznym. Wartość gęstości empirycznej czasu życia w wieku  $x$  oznaczona  $\hat{f}(x)$  jest równa następującemu ilorazowi:

$$\hat{f}(x) = \frac{d_x}{l_0} \quad \text{dla } x = 0, 1, 2, \dots, 100 \quad (16)$$

gdzie:

$d_x$  jest to liczba zgonów w wieku  $x$  (czyli w przedziale  $\langle 0, x+1 \rangle$ ), natomiast  $l_0$  początkowa liczebność badanej populacji. Jeżeli w wieku  $x$  zmarło  $d_x$  osób, to znaczy, że spośród 100 000 osób  $d_x$  przeżyło  $x$  lat. Częstość przeżycia  $x$  lat można więc policzyć ze wzoru (16). W pozostałych punktach przedziału  $(x, x+1)$  wartość funkcji gęstości empirycznej nie jest znana.

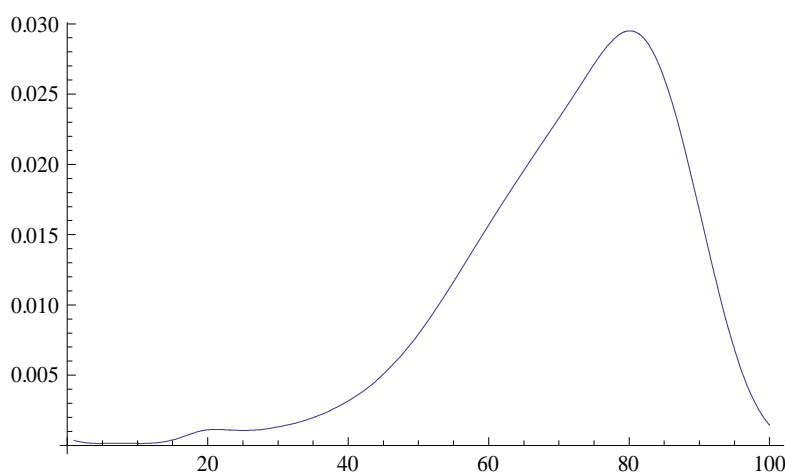
Na rysunku 1 przedstawiony został histogram długości życia mężczyzn polskich w 2009 roku wyznaczony na podstawie badanej próby 100 000 mężczyzn.



Rys. 1. Histogram częstości długości życia

Źródło: Opracowanie własne na podstawie TTŻ mężczyzn w 2009 r.

Na rysunku 2 przedstawiony został natomiast diagram długości życia.



Rys. 2. Diagram częstości długości życia wygładzony numerycznie

Źródło: Opracowanie własne na podstawie TTŻ mężczyzn w 2009 r.

Po wygładzeniu diagramu otrzymuje się funkcję ciągłą.

### 3. BADANIE ZGODNOŚCI ROZKŁADU EMPIRYCZNEGO I TEORETYCZNEGO

Sprawdzimy teraz, czy badana próba pochodzi z populacji, w której czas życia opisany jest za pomocą „prawa życia” Weibulla, czyli czy istnieje zgodność między rozkładem przedstawionym na rysunku 2 i rozkładem teoretycznym opisanym za pomocą funkcji gęstości określonej wzorem (11).

Zadanie rozwiążemy, wykorzystując test  $\lambda$ -Kolmogorowa. Parametry rozkładu Weibulla estymowane będą metodą największej wiarygodności [3, 10].

Estymatory  $\hat{\alpha}$ ,  $\hat{\beta}$  parametrów rozkładu Weibula otrzymane metodą największej wiarygodności (MNW) spełniają równania [6]:

$$\frac{\overline{x^{\hat{\alpha} \ln x}}}{\overline{x^{\hat{\alpha}}}} - \overline{\ln x} = \frac{1}{\hat{\alpha}}$$

$$\hat{\beta} = (\overline{x^{\hat{\alpha}}})^{1/\hat{\alpha}}$$
(17)

gdzie:

$\overline{g(x)}$  oznacza średnią funkcji  $g(x)$ .

Równania te można rozwiązać tylko numerycznie [6, 7].

W przypadku badanej populacji otrzymano następujące wartości estymatorów rozkładu teoretycznego (obliczenia wykonano za pomocą programu z pakietu Matematica):

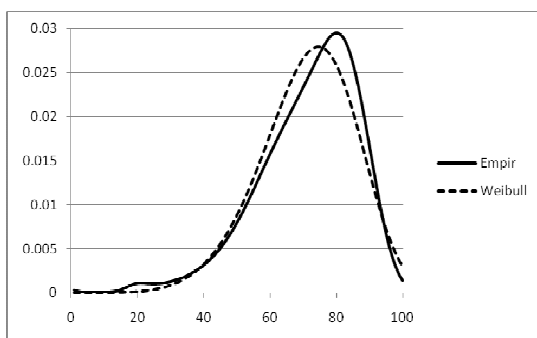
$$\hat{\alpha} = 5,75974$$

$$\hat{\beta} = 77,1482$$

Teoretyczna funkcja gęstości ma więc postać:

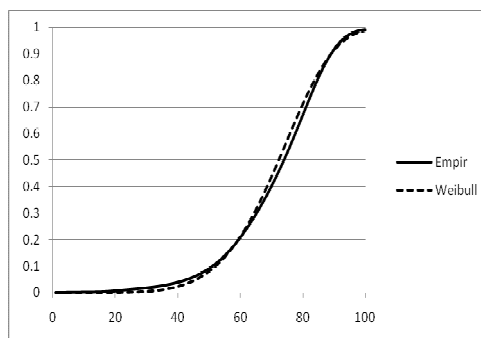
$$f(x) = \begin{cases} 0 & x \leq 0 \\ 0,075 \left(\frac{x}{77,15}\right) \exp\left(-\left(\frac{x}{77,15}\right)^{5,76}\right) & x > 0 \end{cases}$$
(18)

Na rysunku 3 przedstawiono wykresy teoretycznej i empirycznej funkcji gęstości, natomiast na rysunku 4 wykresy dystrybuant.



Rys. 3. Teoretyczna i empiryczna funkcja gęstości

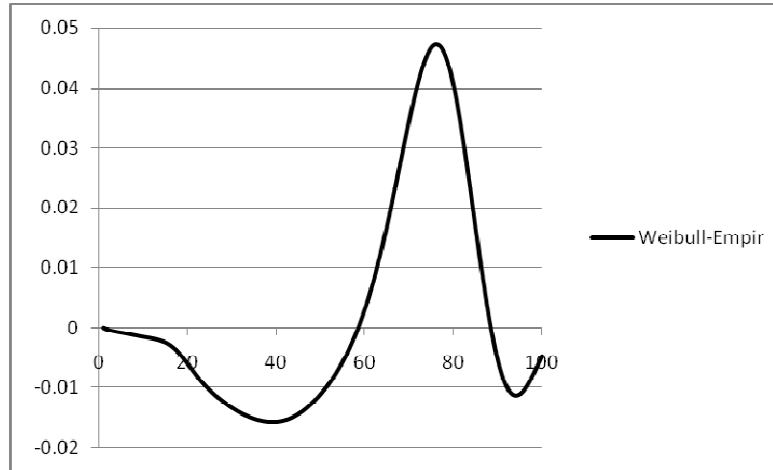
Źródło: Opracowanie własne



Rys. 4. Dystrybuanta rozkładu empirycznego i teoretycznego

Źródło: Opracowanie własne

Funkcja gęstości empiryczna i teoretyczna różnią się we wszystkich przedziałach wieku. Jeszcze bardziej widoczne są różnice między dystrybuantą empiryczną i teoretyczną przedstawioną na rysunku 4. Aby ocenić statystycznie istotność tych różnic dla wszystkich przedziałów wieku, policzono różnicę między wartościami dystrybuanty empirycznej i teoretycznej. Wykres różnic przedstawiony jest na rysunku 5.



Rys. 5. Różnice między wartościami dystrybuanty empirycznej i teoretycznej

Źródło: Opracowanie własne

Jak widać różnice te są dość znaczne. Maksymalna różnica występuje w wieku  $x = 78$  i wynosi około 0,048.

$$\max_x |\hat{F}_n(x) - F(x)| = 0,048$$

gdzie:

$\hat{F}_n(x)$  oznacza wartość dystrybuanty empirycznej w punkcie  $x$ , natomiast  $F(x)$  wartość dystrybuanty teoretycznej w punkcie  $x$ .

Sprawdźmy teraz czy różnice te są na tyle duże, że hipoteza mówiąca o tym, iż próba statystyczna pochodzi z populacji o rozkładzie Weibulla powinna być odrzucona. Do weryfikacji hipotezy o zgodności rozkładu teoretycznego i empirycznego wykorzystany zostanie test  $\lambda$  – Kołmogorowa.

Statystyka testowa  $\lambda$  dla tego testu ma postać:

$$\lambda = \sqrt{n} \max_x |\hat{F}_n(x) - F(x)| \quad (19)$$

Przy założeniu prawdziwości sprawdzanej hipotezy statystyka ta ma znany rozkład który jest stabilizowany zarówno dla małej, jak i dużej próby [10]. Test Kołmogorowa jest testem prawostronnym, czyli hipotezę o jednakowości rozkładu empirycznego i teoretycznego odrzucamy, gdy wartość statystyki testowej przekracza wartość krytyczną. W rozpatrywanym przypadku wartość empiryczna statystyki  $\lambda$  – Kołmogorowa jest równa 15,17 natomiast wartość krytyczna  $\lambda_\alpha$  odczytana z tablic, odpowiadająca poziomowi istotności  $\alpha = 0,05$ , jest równa 1,36. Stąd wynika, że hipotezę o zgodności rozkładu empirycznego i teoretycznego należy odrzucić. Oznacza to, że czas życia w populacji mężczyzn polskich w 2009 roku nie ma rozkładu Weibulla.

Innym rozkładem wykorzystywanym do opisu czasu życia jest rozkład Gomperta, którego funkcja gęstości określona jest wzorem (6).



Na rysunku 6 przedstawiono wykres gęstości empirycznej i gęstości rozkładu Gompertza, którego parametry wyznaczono metodą największej wiarygodności.

Estymatory  $\hat{b}, \hat{\gamma}$  parametrów funkcji gęstości rozkładu Gompertza spełniają następujące równania [7]:

$$\hat{b} = \frac{\hat{\gamma}}{\exp(\hat{\gamma}) - 1} \quad (20)$$

$$\hat{\gamma} \overline{x(\exp(\hat{\gamma}) - 1)} = \overline{\exp(\hat{\gamma})}(\hat{\gamma}^2 - 1) + 1 \quad (21)$$

Równania te rozwiązano numerycznie za pomocą programu z pakietu Mathematica, używając następujące wartości estymatorów:

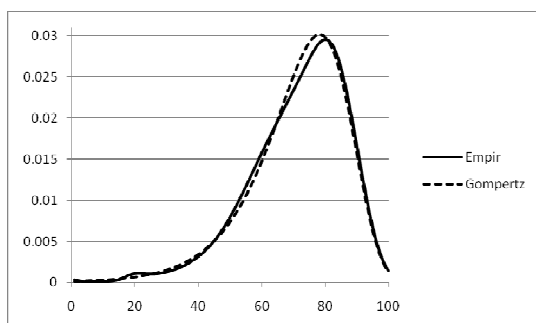
$$\hat{b} = 0,000136$$

$$\hat{\gamma} = 0,08172$$

Funkcja gęstości prawa Gompertza ma więc postać:

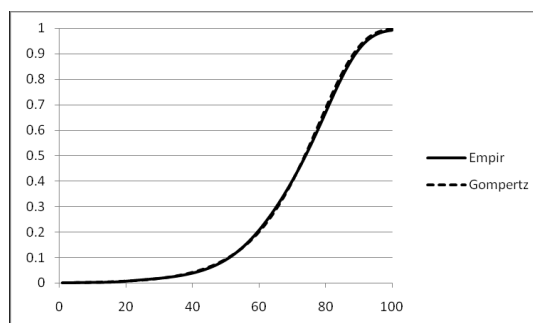
$$f(t) = \exp\{-0,00166 \exp((0,082t) - 1) \exp(0,082t)\} \quad (22)$$

Na rysunku 6 przedstawiono gęstość empiryczną i gęstość rozkładu Gompertza, natomiast na rysunku 7 przedstawiono dystrybuanty tych rozkładów.



Rys. 6. Funkcja gęstości rozkładu empirycznego i rozkładu Gompertza

*Źródło: Opracowanie własne*

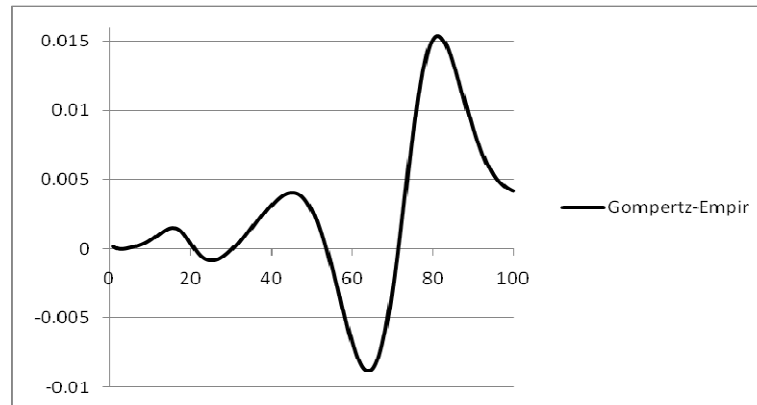


Rys. 7. Dystrybuanta empiryczna i teoretyczna

*Źródło: Opracowanie własne*

Widać, że zgodność gęstości empirycznej i teoretycznej jest bardzo duża dla populacji w wieku powyżej 80 lat. Fakt ten został dawno zauważony przez demografów i aktuariuszy [1]. W pozostałych grupach wieku zgodność jest znacznie mniejsza. Rozbieżności między rozkładem empirycznym i teoretycznym są znacznie mniej widoczne na wykresach dystrybuant.

Sprawdzimy teraz za pomocą testu statystycznego  $\lambda$  – Kołmogorowa istotność różnic między tymi funkcjami. W tym celu policzymy wartości  $\hat{F}_n(x) - F(x)$  dla wszystkich wartości  $x$  i przedstawimy je na rysunku 8.



Rys. 8. Różnice między dystrybucją empiryczną i dystrybucją rozkładu Gompertza

Źródło: Opracowanie własne

Jak można odczytać z wykresu, maksymalna wartość różnicy wynosi około 0,015. Stąd wartość statystyki testowej  $\lambda_e$  jest równa:  $\lambda_e = \sqrt{100000} \cdot 0,015 = 4,73$ . Wartość ta jest większa od wartości krytycznej, która przy poziomie istotności  $\alpha = 0,05$  wynosi 1,96. Hipotezę o tym, że czas życia populacji można opisać rozkładem Gompertza należy odrzucić.

Dalszych prób aproksymacji gęstości empirycznej gęstością teoretyczną nie przeprowadzono. Uznano, że badana populacja nie jest jednorodna, a w takim przypadku na pewno nie znajdzie się rozkładu, który opisywałby rozkład empiryczny. Niejednorodność populacji oznacza, że populacja składa się z pewnych grup jednostek [19]. Wszystkie jednostki w grupie charakteryzują się takim samym rozkładem trwania życia. Jednostki te zostały wymieszane i nie można odróżnić, która jednostka należała do której grupy. Powstał nowy rozkład (mieszanka), którego dystrybuanta oznaczona jest  $G(x)$ . Dystrybuanta  $G(x)$  jest mieszanką rozkładów o dystrybuantach  $G_i(x)$  z poszczególnych grup, przy czym udział rozkładów poszczególnych grup w mieszance, jest proporcjonalny do liczebności tych grup. Liczebności te nie są znane.

Dystrybuanta  $G(x)$  ma postać:

$$G(x) = \sum_{i=1}^m w_i G_i(x) \quad (23)$$

gdzie:

$m$  - liczba grup a  $w_i$  udział poszczególnych grup w mieszance.

Nieznane parametry rozkładu  $G(x)$  estymowane będą metodą minimum chi- kwadrat [7].

Statystyka chi-kwadrat, która będzie minimalizowana, ma następującą postać:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - n\pi_i)^2}{n\pi_i} \quad (24)$$

gdzie:

$k$  oznacza liczbę przedziałów, na które zostały pogrupowane wartości zmiennej losowej,  $\pi_i$  oznacza prawdopodobieństwo, że wartość zmiennej o dystrybuancie  $G(x)$  należy do przedziału  $i$ -tego,  $n_i$  zaobserwowana liczba obserwacji należących do przedziału  $i$ -tego.

Rozpatrzmy najprostszy przypadek, gdy populacja składa się z dwóch grup, rozkład długości życia w pierwszej grupie opisany jest za pomocą rozkładu Weibulla (dystrybuanta określona wzorem (10), natomiast w grupie drugiej za pomocą rozkładu Gompertza (dystrybuanta określona wzorem (7)).

Rozkład mieszanki określony jest za pomocą następującej dystrybuanty:

$$G(x) = w_1(1 - \exp(-\frac{b}{\gamma}(e^{\gamma x} - 1))) + w_2(1 - \exp(-(\frac{x}{\beta})^\alpha)) \quad \text{dla } x > 0 \quad (25)$$

Rozkład  $G(x)$  zależy od pięciu parametrów  $w_1, b, \gamma, \beta, \alpha$ . Udział  $w_2 = 1 - w_1$ . Statystyka chi-kwadrat, która będzie minimalizowana w rozpatrywanym przypadku ma następującą postać:

$$\chi^2 = \sum_{i=0}^{99} \frac{(d_i - l_0 \pi_i)^2}{n \pi_i} \quad (26)$$

gdzie:

$d_i$  jest to obserwowana liczba zgonów w  $i$ -tym przedziale wieku odczytana z TTŻ,  $l_0$  początkowa liczebność populacji, czyli 100 000, natomiast

$$l_0 \pi_i = l_0 (G(i+1) - G(i)) \quad (27)$$

Zadanie minimalizacji statystyki  $\chi^2$  polega na znalezieniu takich wartości parametrów rozkładu  $G(x)$ , aby funkcja określona wzorem (26) miała wartość jak najmniejszą. W rozpatrywanym przypadku funkcja ta zależy od pięciu parametrów (zmiennych), a więc należy znaleźć ekstremum funkcji pięciu zmiennych. Zadanie to zostało rozwiązane numerycznie za pomocą programu z pakietu Mathematica.

Wartości parametrów minimalizujące wartość statystyki  $\chi^2$  są następujące:

$$\hat{\alpha} = 6,124$$

$$\hat{\beta} = 59,308$$

$$\hat{b} = 0,000128$$

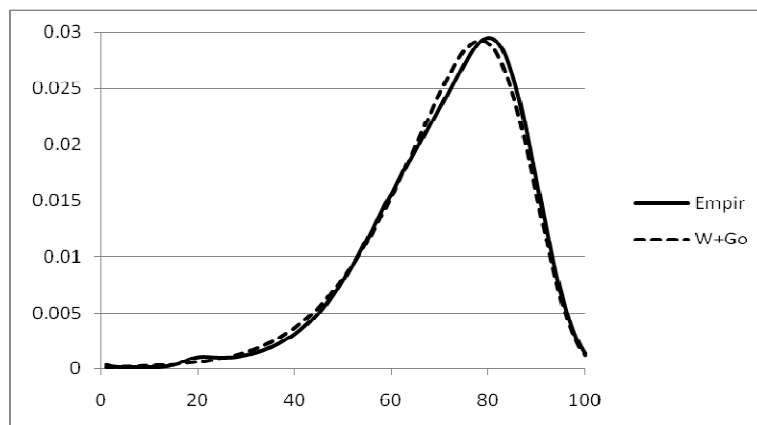
$$\hat{\gamma} = 0,0824$$

$$w_1 = 0,9603$$

$$w_2 = 0,0397$$

Minimalna wartość statystyki  $\chi^2$  jest równa:  $\chi^2 = 419,204$

Wykresy gęstości empirycznej i teoretycznej (mieszanki rozkładu Gompertza i rozkładu Weibulla) przedstawione są na rysunku 9.



Rys. 9. Wykres gęstości empirycznej i gęstości mieszanki

Źródło: Opracowanie własne

Statystyka  $\chi^2$  wykorzystania zostanie do weryfikacji hipotezy o zgodności rozkładu teoretycznego określonego następująco:

$$G(x) = 0,9603(1 - \exp(-0,00155(e^{0,0824 \cdot x} - 1))) + 0,0397(1 - \exp(-\left(\frac{x}{59,308}\right)^{6,124})) \quad (28)$$

Jeżeli próba pochodzi z populacji o rozkładzie  $G(x)$ , to statystyka  $\chi^2$  ma rozkład chi-kwadrat z 94 stopniami swobody [3, 10]. Zbiór krytyczny jest prawostronny, czyli hipotezę zerową odrzucimy, gdy wartość statystyki testowej przekroczy pewną wartość krytyczną zależną od poziomu istotności. Przy poziomie istotności  $\alpha = 0,05$  wartość krytyczna odczytana z tablic rozkładu chi-kwadrat jest równa 117,63 [13]. Otrzymana empiryczna wartość statystyki  $\chi^2$  jest równa 419,24, a więc znajduje się w zbiorze krytycznym. Oznacza to, że na poziomie istotności  $\alpha = 0,05$  hipotezę o tym, że czas trwania życia populacji mężczyzn jest mieszanką 96% jednostek, których czas życia ma rozkład Gompertza i 4% jednostek, których czas życia ma rozkład Weibulla należy odrzucić.

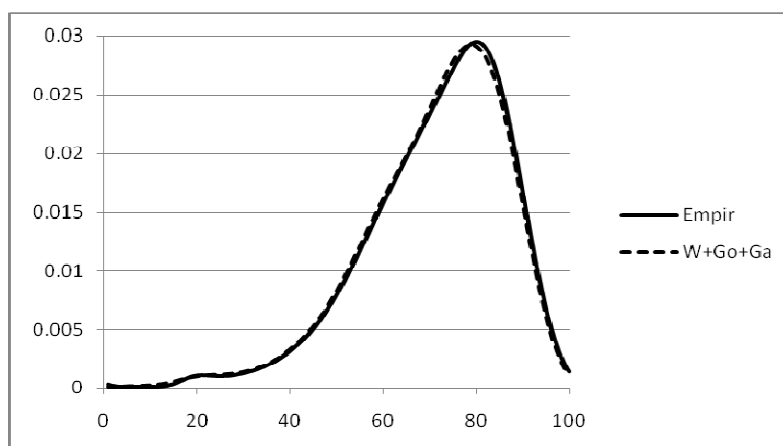
Powstaje podejrzenie, że w populacji są jednostki, których czas życia opisany jest innym niż poprzednio zmieszane rozkłady. Rozpatrzmy teraz mieszankę trzech rozkładów Weibulla, Gompertza i rozkładu gamma. Estymatory parametrów mieszanki otrzymane metodą minimum chi- kwadrat są następujące:

parametry rozkładu Weibulla	$\hat{\alpha} = 8,19,$	$\hat{\beta} = 59,05$
parametry rozkładu Gompertza	$\hat{b} = 0,0001,$	$\hat{\gamma} = 0,08544$
parametry rozkładu gamma	$\hat{a} = 18,826,$	$\hat{c} = 1,164$
waga, z jaką wchodzi do mieszanki rozkład Weibulla	$w_w = 0,0638$	
waga z jaką wchodzi do mieszanki rozkład gamma	$w_\gamma = 0,006148$	
waga, z jaką wchodzi do mieszanki rozkład Gompertza	$w_G = 0,93$	

Wartość statystyki chi-kwadrat jest równa  $\chi^2 = 133,731$

Z przeprowadzonej estymacji wynika, że w analizowanej populacji 93% populacji charakteryzuje się długością życia podlegającą rozkładowi Gompertza, 0,6 % populacji ma rozkład trwania życia opisany rozkładem gamma i 6% populacji charakteryzuje się trwaniem życia opisanym za pomocą rozkładu Weibulla. Ponieważ estymowanych było 8 parametrów rozkładu mieszanki, to statystyka  $\chi^2$  ma rozkład chi-kwadrat z 91 stopniami swobody. Wartość krytyczna przy poziomie istotności  $\alpha = 0,05$  jest równa 117,632, a przy poziomie istotności  $\alpha = 0,02$  jest równa 120,801

Wykres funkcji częstości rozkładu empirycznego i gęstości rozkładu mieszanki przedstawiono na rysunku 10.



Rys. 10. Rozkład empiryczny i rozkład mieszanki trzech rozkładów

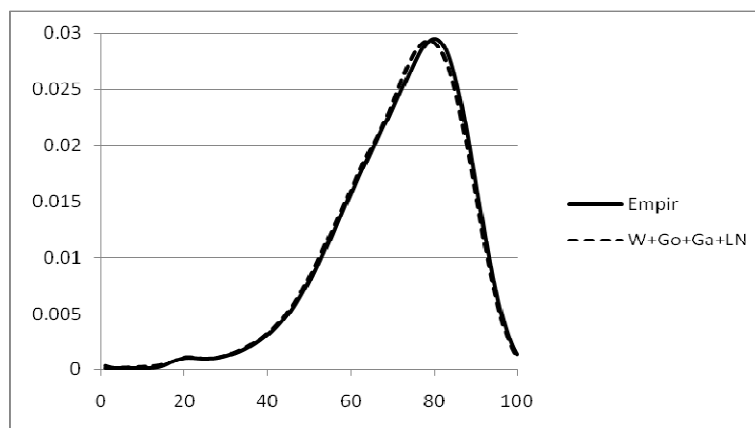
*Źródło: Opracowanie własne*

Wydaje się, że rozkład teoretyczny (mieszanka) zaznaczony na rysunku linią przerywaną i rozkład empiryczny zaznaczony linią ciągłą przebiegają niemal identycznie. Jeśli jednak przeprowadzimy weryfikację statystyczną hipotezy o zgodności tych rozkładów to, okaże się, że zarówno na poziomie istotności  $\alpha = 0,05$ , jak też na poziomie istotności  $\alpha = 0,02$  hipotezę tę trzeba odrzucić, gdyż wartość statystyki testowej, która równa jest 133,71, w obu przypadkach znajduje się w zbiorze krytycznym. Oznacza to, że w skład mieszanki rozkładu teoretycznego wchodzi jeszcze inne rozkłady. Spróbujmy zwiększyć składniki mieszanki o rozkład lognormalny. Jest to rozkład nieujemny zależny od dwóch parametrów. Funkcja gęstości tego rozkładu określona jest wzorem (15) [3]. W tym przypadku rozkład teoretyczny będzie zależał od 11 parametrów.

Estymatory tych parametrów wyznaczone według kryterium minimalnego  $\chi^2$  są następujące:

Estymatory parametrów rozkładu Weibulla	$\hat{\alpha} = 9,505$	$\hat{\beta} = 56,99$
Estymatory parametrów rozkładu Gompertza	$\hat{b} = 0,00011$	$\hat{\gamma} = 0,084$
Estymatory parametrów rozkładu gamma	$\hat{a} = 46,77$	$\hat{c} = 0,439$
Estymatory parametrów rozkładu lognormalnego	$\hat{m} = 4,156$	$\hat{\sigma} = 0,046$
Waga, z jaką wchodzi rozkład Weibulla	$w_W = 0,044$	

Waga, z jaką wchodzi rozkład gamma	$w_\gamma = 0,0044$
Waga, z jaką wchodzi rozkład lognormalny	$w_{LN} = 0,00747$
Waga, z jaką wchodzi rozkład Gompertza	$w_G = 0,944$
Wartość statystyki	$\chi^2 = 108,846$



Rys. 11. Gęstość empiryczna i gęstość mieszanki czterech rozkładów

Źródło: Opracowanie własne

W ostatnim rozpatrywanym przypadku statystyka  $\chi^2$  ma rozkład chi-kwadrat z 88 stopniami swobody. Wartość krytyczna odczytana z tablic przy poziomie istotności  $\alpha = 0,05$  równa jest  $\chi_\alpha^2 = 110,898$ , natomiast przy poziomie istotności  $\alpha = 0,02$  wartość ta równa jest  $\chi_\alpha^2 = 117,340$ . Przy obu poziomach istotności wartość empiryczna znajduje się w zbiorze przyjęcia sprawdzanej hipotezy. Tak więc możemy stwierdzić, że rozkład czasu życia mężczyzn w Polsce w roku 2009 jest mieszanką czterech nieujemnych rozkładów zmiennych losowych. Świadczy to o tym, że populacja jest bardzo silnie zróżnicowana. Ale też z udziałów poszczególnych rozkładów w mieszance wynika, że największą podgrupą w tej populacji bo stanowiącą aż 94,4%, są mężczyźni, których czas życia opisany jest rozkładem Gompertza. Mimo iż jest to grupa bardzo liczna to jednak jej rozkład nie reprezentuje całej populacji.

## LITERATURA

- [1] Balicki A., *Analiza przeżycia i tablice wymieralności*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2006.
- [2] Bowers N, Gerber H., Hickman J., Jones D., Nesbit D. *Actuarial Mathematics*, Itasca: The Society of Actuaries 1986.
- [3] Fisz M., *Rachunek prawdopodobieństwa i statystyka matematyczna*, PWN, Warszawa 1976.
- [4] Gerber H. U., *Life Insurance Mathematics*, Springer-Verlag, 1990.
- [5] Holzer J., *Demografia*, PWE. Warszawa 1999.
- [6] Law A. M., Kelton W., *Simulation Modelling and Analysis*, McGraw-Hill, 1982.

- [7] *Metody oceny i porządkowania ryzyka w ubezpieczeniach życiowych*, pod red. Ostasiewicz S., Wydawnictwo Akademii Ekonomicznej, Wrocław 2000.
- [8] Ostasiewicz S., *Ocena umieralności w populacjach o zróżnicowanej witalności*, [w:] „Roczniki Kolegium Analiz Ekonomicznych”, Zeszyt 21/2010.
- [9] Ostasiewicz S., *Składki w wybranych typach ubezpieczeń życiowych*, Wydawnictwo Akademii Ekonomicznej, Wrocław 2000.
- [10] Ostasiewicz W., *Propedeutyka probabilistyki*, Wydawnictwo Akademii Ekonomicznej, Wrocław 2000.
- [11] Rao C., *Modele liniowe statystyki matematycznej*, PWN, Warszawa, 1982.
- [12] [online]. [dostęp: 2011]. Dostępny w Internecie: [http://www.stat.gov.pl/cps/rde/xbcr/gus/PUBL\\_lud\\_trwanie\\_zycia\\_2009.pdf](http://www.stat.gov.pl/cps/rde/xbcr/gus/PUBL_lud_trwanie_zycia_2009.pdf).
- [13] Trwanie życia w 2009 r. Life expectancy tables of Poland. Informacje i opracowania statystyczne. GUS. Warszawa 2011.
- [14] [online]. [dostęp: 2011]. Dostępny w Internecie: [http://pl.wikisource.org/wiki/Tabela\\_rozk%C5%82adu\\_chi-kwadrat](http://pl.wikisource.org/wiki/Tabela_rozk%C5%82adu_chi-kwadrat)

## APPROXIMATION OF SURVIVAL FUNCTION FOR HETEROGENEOUS POPULATIONS

### Summary

*For a long time demographers and actuaries have been deliberating the issue of the laws of life. A number of proposed survival functions turned out to be unsatisfactory when they were applied empirically. One of the ways to overcome the difficulties is to modify the general survival functions by introducing an additional formula characterizing the frailty of individuals. Another way is to use a mixture of appropriate distributions. In this contribution the latter approach to determine the survival time of men in the Polish population in 2009 is applied.*

**Key words:** *lifetime model, death rate, men, mixture of distributions, heterogeneity*

Tabela 1. Tabela trwania życia 2009

56

TABL. A. TABLICA TRWANIA ŻYCIA 2009 LIFE TABLE OF POLAND 2009		OGÓŁEM TOTAL		MĘŻCZYŹNI MALES		
Wiek Age	Liczba dożywających Number of survivors	Prawdopodobieństwo zgonu Probability of dying	Liczba zmarłych Number deceased	Ludność stacjonarna Stationary population		Przeciętne dalsze trwanie życia Life expectancy
				w wieku x at age x	skumulowana cumulated	
x	$l_x$	$q_x$	$d_x$	$L_x$	$T_x$	$e_x$
0	100000	0,00604	604	99461	7153083	71,53
1	99396	0,00035	35	99379	7053622	70,96
2	99362	0,00024	24	99350	6954243	69,99
3	99338	0,00017	17	99330	6854893	69,01
4	99321	0,00015	14	99314	6755563	68,02
5	99307	0,00014	14	99300	6656249	67,03
6	99293	0,00015	15	99285	6556950	66,04
7	99278	0,00016	15	99270	6457665	65,05
8	99262	0,00015	15	99254	6358395	64,06
9	99247	0,00015	15	99240	6259141	63,07
10	99232	0,00014	14	99225	6159901	62,08
11	99218	0,00014	14	99211	6060676	61,08
12	99204	0,00016	16	99196	5961464	60,09
13	99188	0,00020	20	99178	5862268	59,10
14	99168	0,00027	27	99155	5763090	58,11
15	99141	0,00038	38	99122	5663935	57,13
16	99104	0,00053	53	99077	5564812	56,15
17	99051	0,00071	70	99015	5465735	55,18
18	98980	0,00089	88	98936	5366720	54,22
19	98892	0,00103	102	98841	5267784	53,27
20	98790	0,00112	110	98734	5168943	52,32
21	98679	0,00114	113	98623	5070209	51,38
22	98566	0,00113	112	98510	4971586	50,44
23	98455	0,00111	109	98400	4873075	49,50
24	98346	0,00109	107	98292	4774675	48,55
25	98239	0,00108	106	98185	4676383	47,60
26	98132	0,00109	107	98079	4578198	46,65
27	98026	0,00112	109	97971	4480119	45,70
28	97916	0,00117	114	97859	4382148	44,75
29	97802	0,00125	122	97741	4284289	43,81
30	97680	0,00133	130	97615	4186548	42,86
31	97550	0,00143	140	97480	4088933	41,92
32	97411	0,00155	151	97335	3991453	40,98
33	97259	0,00169	164	97177	3894118	40,04
34	97095	0,00185	179	97005	3796941	39,11
35	96915	0,00203	196	96817	3699936	38,18
36	96719	0,00223	215	96611	3603118	37,25
37	96504	0,00245	236	96385	3506507	36,34
38	96267	0,00270	260	96137	3410122	35,42
39	96008	0,00297	285	95865	3313984	34,52
40	95723	0,00327	313	95566	3218119	33,62
41	95410	0,00360	344	95238	3122553	32,73
42	95066	0,00397	378	94877	3027315	31,84
43	94688	0,00438	415	94480	2932439	30,97
44	94273	0,00484	456	94045	2837958	30,10
45	93817	0,00534	501	93566	2743914	29,25
46	93316	0,00589	550	93041	2650348	28,40
47	92766	0,00650	603	92464	2557307	27,57
48	92162	0,00716	660	91832	2464843	26,74
49	91502	0,00788	721	91142	2373011	25,93
50	90781	0,00866	786	90388	2281869	25,14

Trwanie życia w 2009 r.



TABL. A. TABLICA TRWANIA ŻYCIA 2009 (cd.)  
LIFE TABLE OF POLAND 2009 (cont.)OGÓLEM  
TOTAL  
MĘŻCZYŹNI  
MALES

Wiek Age	Liczba dożywających Number of survivors	Prawdopodobieństwo zgonu Probability of dying	Liczba zmarłych Number deceased	Ludność stacjonarna Stationary population		Przeciętne dalsze trwanie życia Life expectancy
				w wieku x at age x	skumulowana cumulated	
x	$l_x$	$q_x$	$d_x$	$L_x$	$T_x$	$e_x$
51	89995	0,00949	854	89568	2191481	24,35
52	89141	0,01038	925	88679	2101913	23,58
53	88216	0,01133	1000	87716	2013234	22,82
54	87216	0,01234	1076	86678	1925518	22,08
55	86140	0,01341	1155	85563	1838840	21,35
56	84985	0,01453	1235	84367	1753277	20,63
57	83750	0,01571	1316	83092	1668910	19,93
58	82434	0,01695	1397	81735	1585818	19,24
59	81037	0,01824	1478	80298	1504083	18,56
60	79559	0,01958	1557	78780	1423785	17,90
61	78001	0,02097	1636	77183	1345005	17,24
62	76365	0,02244	1713	75509	1267821	16,60
63	74652	0,02397	1789	73757	1192313	15,97
64	72863	0,02559	1865	71930	1118555	15,35
65	70998	0,02731	1939	70028	1046625	14,74
66	69059	0,02915	2013	68052	976596	14,14
67	67045	0,03113	2087	66002	908544	13,55
68	64958	0,03327	2161	63878	842543	12,97
69	62797	0,03559	2235	61680	778665	12,40
70	60562	0,03814	2310	59407	716985	11,84
71	58253	0,04094	2385	57060	657578	11,29
72	55868	0,04405	2461	54637	600518	10,75
73	53407	0,04751	2537	52138	545880	10,22
74	50870	0,05138	2614	49563	493742	9,71
75	48256	0,05572	2689	46911	444179	9,20
76	45567	0,06057	2760	44187	397267	8,72
77	42807	0,06596	2823	41395	353080	8,25
78	39984	0,07191	2875	38546	311685	7,80
79	37108	0,07844	2911	35653	273139	7,36
80	34198	0,08555	2925	32735	237486	6,94
81	31272	0,09323	2915	29814	204751	6,55
82	28357	0,10148	2878	26918	174937	6,17
83	25479	0,11030	2810	24074	148019	5,81
84	22669	0,11971	2714	21312	123945	5,47
85	19955	0,12973	2589	18661	102633	5,14
86	17366	0,14041	2438	16147	83972	4,84
87	14928	0,15176	2265	13795	67825	4,54
88	12662	0,16386	2075	11625	54030	4,27
89	10587	0,17676	1871	9652	42405	4,01
90	8716	0,19041	1660	7886	32754	3,76
91	7056	0,20512	1447	6333	24868	3,52
92	5609	0,22063	1237	4990	18535	3,30
93	4371	0,23694	1036	3854	13545	3,10
94	3336	0,25406	847	2912	9691	2,91
95	2488	0,27196	677	2150	6779	2,72
96	1812	0,29063	526	1548	4629	2,56
97	1285	0,31004	398	1086	3081	2,40
98	887	0,33015	293	740	1995	2,25
99	594	0,35093	208	490	1255	2,11
100	385	0,37234	144	314	765	1,99

Life tables of Poland 2009