ROZDZIAŁ 3

Tablice trwania życia

1. Przyszły czas życia

Osobę, która ukończyła x lat życia, będziemy nazywać x-latkiem i oznaczać symbolem (x). Jej **przyszły czas życia**, tzn. od chwili x do chwili śmierci, będziemy oznaczać przez T_x. Wartości T_x są nieujemne, ale nie muszą być całkowite!

Oczywiście T_x dla danej osoby (x) żyjącej nie jest znane. Zakładamy zatem, że T_x jest zmienną losową i że dany jest jej rozkład (dystrybuanta)

$$F_x(t) = \mathrm{P}(T_x \leqslant t), \quad t \geqslant 0.$$

Inaczej, $F_x(t)$ jest prawdopodobieństwem, że x-latek umrze przed upływem czasu t, tzn. przed chwilą $x + t$. Będziemy zawsze zakładać, że F_x ma gęstość f_x, tzn. że dla dowolnych $0 \leqslant a \leqslant b$

$$\mathrm{P}(a \leqslant T_x \leqslant b) = \int_a^b f_x(t)dt,$$

lub równoważnie

$$F_x'(t) = f_x(t)$$

dla prawie wszystkich t.

Przez s_x oznaczymy **funkcję przeżycia**

$$s_x(t) = \mathrm{P}(T_x > t) = \int_x^\infty f_x(t)dt.$$

Będziemy używać następujących oznaczeń:

- prawdopodobieństwo, że (x) umrze przed upływem czasu t

$${}_tq_x = F_x(t);$$

- prawdopodobieństwo, że (x) przeżyje jeszcze t lat

$${}_tp_x = 1 - F_x(t) = s_x(t);$$

  Oczywiście ${}_tq_x + {}_tq_x = 1$.

- prawdopodobieństwo, że (x) przeżyje jeszcze s lat, a następnie umrze w ciągu czasu t

$$\begin{aligned} {}_{s|t}q_x &= \mathrm{P}(s < T_x \leqslant s + t) \\ &= F_x(s + t) - F_x(s) \\ &= {}_{s+t}q_x - {}_sq_x = {}_sp_x - {}_{s+t}p_x; \end{aligned}$$

- prawdopodobieństwo, że (x) przeżyje kolejne t lat, pod warunkiem, że przeżyje najpierw co najmniej s lat
$$\begin{aligned} {}_tp_{[x]+s} &= \mathrm{P}(T_x > s+t \mid T_x > s) = \\ &= \frac{1-F_x(s+t)}{1-F_x(s)} = \frac{{}_{s+t}p_x}{{}_sp_x}, \end{aligned}$$
oraz
- przeciwne prawdopodobieństwo warunkowe
$$\begin{aligned} {}_tq_{[x]+s} &= \mathrm{P}(T_x \leqslant s+t \mid T_x > s) = \\ &= \frac{F_x(s+t)-F_x(s)}{1-F_x(s)} = \frac{{}_{s|t}q_x}{{}_sp_x}. \end{aligned}$$

Uwaga. Jeżeli jakiś indeks jest równy 1, to można go pominąć, np.

$${}_1p_x = p_x, \quad {}_{t|1}q_x = {}_{t|}q_x.$$

Przykład 8. Niech $x = 50$, $t = 5$ oraz $s = 10$. Wtedy:

- $q_x = q_{50}$ oznacza prawdopodobieństwo, że osoba 50 letnia umrze w ciągu kolejnego roku;
- $p_x = p_{50}$ oznacza prawdopodobieństwo, że osoba 50-letnia przeżyje kolejny rok;
- ${}_tq_x = {}_5q_{50}$ oznacza prawdopodobieństwo, że osoba 50–letnia umrze przed osiągnięciem 55 lat;
- ${}_tp_x = {}_5p_{50}$ oznacza prawdopodobieństwo, że osoba 50–letnia dożyje wieku 55 lat;
- ${}_{s|t}q_x = {}_{10|5}q_{50}$ oznacza prawdopodobieństwo, że osoba 50–letnia umrze pomiędzy 60 a 65 rokiem życia;
- ${}_tq_{[x]+s} = {}_5q_{[50]+10}$ oznacza prawdopodobieństwo, że osoba 50–letnia umrze przed osiągnięciem 65 lat, pod warunkiem, że dożyła ona 60 lat.
- ${}_tp_{[x]+s} = {}_5p_{[50]+10}$ oznacza prawdopodobieństwo, że osoba 50–letnia dożyje wieku 65 lat, pod warunkiem, że dożyła ona 60 roku życia;

Uwaga. ${}_tp_{x+s} = {}_5p_{60}$ oznacza prawdopodobieństwo, że osoba 60–letnia dożyje 65 roku życia. Jest to inne prawdopodobieństwo niż ${}_tp_{[x]+s} = {}_5p_{[50]+10}$, chociaż na pierwszy rzut oka mówią one o tym samym: w obydwu przypadkach chodzi o przeżycie od 60 do 65 roku życia. Ale dotyczą one różnych populacji: ${}_5p_{60}$ dotyczy 60-latków, a ${}_5p_{[50]+10}$ dotyczy 50-latków.

Zachodzą następujące równości

$${}_{s+t}p_x = {}_sp_x\,{}_tp_{[x]+s}$$

$${}_{s|t}q_x = {}_sp_x\,{}_tq_{[x]+s},$$

a więc

$$ {}_kp_x = p_x \prod_{i=1}^{k-1} p_{[x]+i}. $$

Wartością oczekiwaną przyszłego czasu życia T_x nazywamy

$$ \mathring{e}_x = ET_x = \int_0^\infty t f_x(t) dt. $$

Zakładamy, że $\mathring{e}_x < \infty$ dla każdego x. Zauważmy, że

$$ f_x(t) = -\frac{d({}_tp_x)}{dt}, $$

a więc całkując przez części

$$ \mathring{e}_x = [-t\, {}_tp_x]_0^\infty + \int_0^\infty {}_tp_x dt. $$

Ostatecznie

$$ \mathring{e}_x = \int_0^\infty {}_tp_x dt. $$

Podobnie

$$ ET_x^2 = \int_0^\infty t^2 f_x(t) dt = 2\int_0^\infty t\, {}_tp_x\, dt, $$

a więc wariancję przyszłego czasu trwania życia można obliczyć ze wzoru

$$ \begin{aligned} \operatorname{Var} T_x &= ET_x^2 - (ET_x)^2 \\ &= 2\int_0^\infty t\, {}_tp_x\, dt - \left(\mathring{e}_x\right)^2. \end{aligned} $$

Natężeniem śmiertelności x-latka w chwili t liczonego od chwili obecnej (tzn. w chwili $x+t$) nazywamy wielkość

$$ \mu_{[x]+t} = \frac{f_x(t)}{1 - F_x(t)}. $$

Rozważmy prawdopodobieństwo warunkowe śmierci (x) w krótkim przedziale czasu $[t, t+h]$ pod warunkiem, że (x) przeżyje do czasu t

$$ {}_hq_{[x]+t} = \mathrm{P}(T_x \leqslant t+h \mid T_x > t) = \frac{F_x(t+h) - F_x(t)}{1 - F_x(t)}. $$

Na mocy twierdzenia o wartości średniej mamy

$$ F_x(t+h) - F_x(t) = hf_x(t + \theta h), $$

dla pewnego $\theta \in [0, 1]$, a więc

$$ \frac{{}_hq_{[x]+t}}{h} = \frac{f_x(t+\theta h)}{1 - F_x(t)}. $$

Jeżeli gęstość jest ciągła w p-cie t, to dostajemy

$$ \lim_{h \to 0^+} \frac{{}_hq_{[x]+t}}{h} = \mu_{[x]+t} $$

Zatem prawdopodobieństwo śmierci x-latka w krótkim przedziale czasu $[t, t+h]$ jest proporcjonalne do długości tego przedziału ze współczynnikiem proporcjonalności $\mu_{[x]+t}$.

Dalej zauważmy, że

$$\mu_{[x]+t} = -\frac{1}{{}_tp_x}\frac{d({}_tp_x)}{dt} = -\frac{d(\log {}_tp_x)}{dt}$$

a więc

$$1 - F_x(t) = \exp\left(-\int_0^t \mu_{[x]+s} ds\right)$$

oraz

$$F_x(t) = \int_0^t {}_sp_x\, \mu_{[x]+s} ds.$$

Inaczej mówiąc natężenie śmiertelności wyznacza rozkład przyszłego czasu życia.

Obciętym czasem życia nazywamy zmienną losową

$$K_x = \lfloor T_x \rfloor,$$

gdzie $\lfloor a \rfloor$ oznacza część całkowitą liczby rzeczywistej a, czyli największą liczbę całkowitą niewiększą niż a. Inaczej $\lfloor a \rfloor = k$ wtedy, i tylko wtedy, gdy

$$k \leqslant a < k + 1.$$

Zatem K_x oznacza liczbę ukończonych przyszłych lat życia x-latka. Funkcja prawdopodobieństwa zmiennej losowej K_x dana jest wzorem

$$\mathrm{P}(K_x = k) = \mathrm{P}(k \leqslant T_x < k + 1) = {}_{k|1}q_x = {}_kp_x\, q_{[x]+k}.$$

Zatem **oczekiwany obcięty przyszły czas życia** jest dany wzorem

$$e_x = \sum_{k=1}^{\infty} k\mathrm{P}(K_x = k) = \sum_{k=1}^{\infty} k\, {}_kp_x\, q_{[x]+k}.$$

K_x jest zmienną losową o wartościach całkowitych nieujemnych, a więc jej wartość oczekiwaną można również obliczyć ze wzoru

$$e_x = \sum_{k=1}^{\infty} \mathrm{P}(K \geqslant k) = \sum_{k=1}^{\infty} {}_kp_x.$$

Wariancja K_x wyraża się wzorem

$$\mathrm{Var}\, K_x = EK_x^2 - (EK_x)^2$$

lub

$$\mathrm{Var}\, K_x = 2\sum_{k=0}^{\infty} k\, {}_kp_t - (e_x)^2\,.$$

Rozważmy pewną populację (grupę ludzi) urodzonych w tym samym roku złożoną z l_0 osób. Zwykle $l_0 = 10000$. Typowa **tablica trwania życia** jest to zbiór par postaci (x, l_x), dla $x = 0, 1, \ldots, 100$, gdzie l_x oznacza oczekiwaną liczbę osób z populacji, które żyją w chwili x. Zauważmy, że dla całkowitych $k \geqslant 0$

$$\frac{l_k}{l_0} = \mathrm{P}(T_0 \geqslant k) = \mathrm{P}(K_0 \geqslant k) = {}_kp_0,$$

a więc z tablic trwania życia można wyznaczyć bezpośrednio rozkład zmiennej losowej K_0, czyli rozkład obciętego przyszłego czasu trwania życia noworodka.

Rozważymy teraz problem jak przy znanych wartościach ${}_kp_0$, $k = 0, 1, 2, \ldots$ wyznaczyć ${}_tp_x$ dla dowolnych wartości $x \geqslant 0$ i $t \geqslant 0$. W tym celu o badanej populacji musimy dokonać pewnych założeń zwanych hipotezami.

Zajmiemy się najpierw problemem wyznaczenia rozkładu K_x gdy znany jest rozkład K_0 lub ogólniej — wyznaczenia rozkładu T_x gdy znany jest rozkład T_0. Problem ten sprowadza się do problemu wyznaczenia prawdopodobieństw ${}_tp_x$ dla dowolnych $x \geqslant 0$, gdy znane są ${}_tp_0$. Problemu tego dotyczą **hipotezy agregacyjne**:

- hipoteza jednorodnej populacji (HJP)
- hipoteza agregacyjna (HA).

Następnie zajmiemy się problemem wyznaczenia wartości ${}_tp_x$, dla dowolnego $t \geqslant 0$, gdy znane są tylko ${}_kp_x$ dla $k = 0, 1, 2 \ldots$. Problemu tego dotyczą **hipotezy interpolacyjne**:

- hipoteza jednostajności (HU);
- hipoteza wykładnicza (HCFM);
- hipoteza Balducciego (HB).

2. Hipotezy agregacyjne

Hipoteza jednorodnej populacji

Rozważmy pewną populację osób (w różnym wieku) i załóżmy, że każda z tych osób w chwili urodzin otrzymała losowy czas życia T_0 o ustalonym, ale jednakowym rozkładzie opisanym funkcja przeżycia

$$s(t) = \mathrm{P}(T_0 > t).$$

Jeżeli spełniony jest warunek

$$\mathrm{P}(T_x > t) = \mathrm{P}(T_0 > x + t \mid T_0 > x)$$

dla wszystkich $x, t \geqslant 0$, to mówimy, że populacja ta spełnia **hipotezę jednorodnej populacji** (HJP). Warunek ten oznacza, że przyszły czas życia T_x osoby, która dożyła wieku x jest taki sam jak rozkład $T_0 - x$ przy warunku $T_0 > x$.

Zauważmy jeszcze, że

$$\mathrm{P}(T_0 > x + t \mid T_0 > x) = \frac{\mathrm{P}(T_0 > x + t)}{\mathrm{P}(T_0 > x)} = \frac{{}_{x+t}p_0}{{}_xp_0},$$

a więc HJP jest równoważna warunkowi

$${}_tp_x = \frac{{}_{x+t}p_0}{{}_xp_0}$$

dla wszystkich $x, t \geqslant 0$. Inaczej mówiąc, przy założeniu HJP, rozkład T_x dla $x \geqslant 0$, wyraża się przez rozkład T_0.

Niech

$$\mu_t = -\frac{s'(t)}{s(t)}, \quad t \geqslant 0,$$

będzie natężeniem zgonów związanym ze zmienną T_0. Wiemy już, że

$$s(t) = \exp\left(-\int_0^t \mu_u\, du\right).$$

TWIERDZENIE 1. *Hipoteza HJP jest równoważna następującym warunkom:*

$$_tp_{[x]+u} = {}_tp_{x+u} \tag{*}$$

lub

$$\mu_{[x]+t} = \mu_{x+t}. \tag{**}$$

DOWÓD. Jeśli zachodzi HJP, to

$$_tp_{x+u} = \frac{_{x+u+t}p_0}{_{x+u}p_0} = \frac{_{x+u+t}p_0 / {}_xp_0}{_{x+u}p_0 / {}_xp_0} = \frac{_{t+u}p_x}{_up_x} = {}_tp_{[x]+u}.$$

W drugą stronę

$$_tp_{[x]+u} = \frac{_{t+u}p_x}{_up_x},$$

a więc jeśli $_tp_{[x]+u} = {}_tp_{x+u}$, to

$$_tp_{x+u} = \frac{_{t+u}p_x}{_up_x}.$$

Kładąc $x = 0$ otrzymujemy HJP. Zatem warunek $(*)$ jest równoważny HJP.

Dalej mamy

$$\mu_{[x]+t} = -\frac{1}{_tp_x}\frac{d(_tp_x)}{dt}$$

a więc jeżeli zachodzi HJP, to

$$\mu_{[x]+t} = \frac{_xp_0}{_{x+t}p_0}\frac{d}{dt}\left(\frac{_{x+t}p_0}{_xp_0}\right) = -\frac{1}{_{x+t}p_0}\frac{d(_{x+t}p_0)}{dt} = \mu_{[0]+x+t} = \mu_{x+t}.$$

Z drugiej strony, jeżeli $\mu_{[x]+t} = \mu_{x+t}$, to

$$\begin{aligned} _tp_x &= \exp\left(-\int_0^t \mu_{[x]+u}du\right) \\ &= \exp\left(-\int_0^t \mu_{x+u}du\right) \\ &= \exp\left(-\int_x^{x+t} \mu_u du\right) = \frac{_{x+t}p_0}{_xp_0}. \end{aligned}$$

Zatem $(**)$ jest również równoważny HJP. □

WNIOSEK 1. *Jeżeli zachodzi HJP, to*

$$_tp_x = \exp\left(-\int_x^{x+t} \mu_u du\right)$$

oraz

$$\mathring{e}_x = \frac{1}{s(x)}\int_x^\infty s(y)dy.$$

Dowód. Pierwszą równość wykazaliśmy w dowodzie Tw. 1. Druga równość wynika z następujących przekształceń

$$\mathring{e}_x = \int_0^\infty {}_tp_x dt = \int_0^\infty \frac{s(x+t)}{s(x)} dt = \frac{1}{s(x)} \int_0^\infty s(x+t) dt.$$

$\square$

Przykłady rozkładów zmiennej T_0

- Rozkład de Moivre'a (1729), który postulował istnienie maksymalnego wieku jednostki $\omega = 100$ lat. Rozkład T_0 miał być jednostajny na przedziale $[0, \omega]$, a więc
$$s(t) = 1 - \frac{t}{\omega}, \quad 0 \leqslant t \leqslant \omega,$$
oraz
$$\mu_t = \frac{1}{\omega - t}, \quad 0 \leqslant t \leqslant \omega.$$
Przy dodatkowym założeniu HJP rozkład T_x jest rozkładem jednostajnym na $[0, \omega - x]$, a więc
$${}_tp_x = 1 - \frac{t}{\omega - x}.$$
- Rozkład Gompertza (1824), który postulował, że natężenie zgonów jest wykładnicze postaci
$$\mu_t = Bc^t, \quad t > 0,$$
gdzie $B > 0$ i $c > 1$.
- Rozkład Makehama (1860), który zaproponował, że
$$\mu_t = A + Bc^t, \quad t > 0,$$
gdzie $B > 0$ i $c > 1$ oraz $A \geqslant -B$.
- Rozkład Weibulla (1939), który zakładał, że
$$\mu_t = kt^n, \quad t \geqslant 0,$$
gdzie $k > 0$, $n > 0$.

Hipoteza agregacji.

Przypomnijmy, że jeżeli T_x oznacza przyszły czas trwania życia x-latka, to $K_x = \lfloor T_x \rfloor$ oznacza liczbę pełnych lat, które przeżyje jeszcze x-latek. Zatem oczywiście dla $k \in \mathbb{N}$

$$\mathrm{P}(K_x \geqslant k) = \mathrm{P}(T_x \geqslant k).$$

Powiemy, że dla danej populacji spełniona jest **hipoteza agragacji** (HA), jeżeli dla dowolnych $x, k \in \mathbb{N}$ zachodzi zależność

$$\mathrm{P}(K_x \geqslant k) = \mathrm{P}(K_0 \geqslant x + k \mid K_0 \geqslant x). \tag{HA}$$

Korzystając z powyższej równości łatwo pokazać, że jeżeli dana populacja spełnia HJP, to spełnia również HA, chociaż na odwrót tak być nie musi. Rownoważnie zamiast zależności (HA) można zakładać warunek

$$\mathrm{P}(K_x = k) = \mathrm{P}(K_0 = x + k \mid K_0 \geqslant x). \tag{HA'}$$

Twierdzenie 2. *Hipoteza HA jest równoważna każdemu z następujących warunków:*

(a) ${}_kp_{[0]+x} = {}_kp_x$;
(b) ${}_kq_{[0]+x} = {}_kq_x$;
(c) $p_{[x]+k} = p_{x+k}$;
(d) $p_{[x]+k} = p_{x+k}$.

Dowód. Oczywiście (HA) $\Leftrightarrow$ (a) i (HA') $\Leftrightarrow$ (b). Ponadto (a) $\Leftrightarrow$ (b) i (c) $\Leftrightarrow$ (d). Wystarczy zatem udowodnić, że (a) $\Rightarrow$ (c) $\Rightarrow$ (HA).

Załóżmy zatem, że zachodzi (a). Wtedy

$$\begin{aligned} p_{[x]+k} &= \frac{{}_{k+1}p_x}{{}_kp_x} = \frac{{}_{k+1}p_{[0]+x}}{{}_kp_{[0]+x}} = \frac{\mathrm{P}(K_0 \geqslant x + k + 1 \mid K_0 \geqslant x)}{\mathrm{P}(K_0 \geqslant x + k \mid K_0 \geqslant x)} \\ &= \frac{\mathrm{P}(K_0 \geqslant x + k + 1)}{\mathrm{P}(K_0 \geqslant x + k)} = \mathrm{P}(K_0 \geqslant x + k + 1 \mid K_0 \geqslant x + k) = p_{x+k}, \end{aligned}$$

a więc (a) $\Rightarrow$ (c).

Jeżeli zachodzi (c), to

$$\begin{aligned} \mathrm{P}(K_x \geqslant k) = {}_kp_x = p_x \prod_{i=1}^{k-1} p_{[x]+i} &= p_{[0]+x} \prod_{i=1}^{k-1} p_{[0]+x+i} \\ &= \frac{p_0 \prod_{i=1}^{x+k-1} p_{[0]+i}}{p_0 \prod_{i=1}^{x-1} p_{[0]+i}} \\ &= \mathrm{P}(K_0 \geqslant x + k \mid K_0 \geqslant x). \end{aligned}$$

Zatem (c) $\Rightarrow$ (HA). □

Przypomnijmy teraz, że dla dowolnych $s, t, x \geqslant 0$ mamy

$${}_{s+t}p_x = {}_sp_x\, {}_tp_{[x]+s}.$$

Wniosek 2. *Jeżeli zachodzi HA, to dla* $k, n, x \in \mathbb{N}$

$${}_{k+n}p_x = {}_kp_x \cdot {}_np_{x+k} = {}_np_x \cdot {}_kp_{x+n}$$

oraz

$${}_kp_x = p_x p_{x+1} \dots p_{x+k-1}.$$

Mamy $\mathrm{P}(K_x = k) = {}_kp_x q_{[x]+k}$, a więc jeśli dana jest tablica liczb p_x, $x = 1, 2, \dots$, to przy założeniu HA możemy wyznaczyć rozkład K_x dla każdego $x = 1, 2, \dots$ (bo $q_{[x]+k} = q_{x+k} = 1 - p_{x+k}$)

PRZYKŁAD 9. Prawdopodobieństwo, że obcięty czas życia 50-latka wynosi 60 jest równe

$$P(K_{50} = 10) = {}_{10}p_{50}(1 - p_{60}) = p_{50}p_{51} \dots p_{59}(1 - p_{60}).$$

TWIERDZENIE 3. *Przy założeniu HA dla każdego $x = 0, 1, 2, \dots$ zachodzi wzór*

$$e_x = \frac{1}{{}_xp_0} \sum_{k=x+1} {}_kp_0.$$

DOWÓD. Mamy

$$\begin{aligned} e_x &= \sum_{k=1}^{\infty} {}_kp_x = \sum_{k=1}^{\infty} P(K_x \geqslant k) \\ &= \frac{1}{P(K_0 \geqslant x)} \sum_{k=1}^{\infty} P(K_0 \geqslant x + k) = \frac{1}{{}_xp_0} \sum_{k=1}^{\infty} {}_{x+k}p_0 = \frac{1}{{}_xp_0} \sum_{k=x+1} {}_kp_0. \end{aligned}$$

□

Hipotezy HJP i HA nie zawsze muszą być spełnione. Jeśli bowiem zachodzi np. HA, to na mocy powyższego twierdzenia mamy na przykład

$$p_{[50]+1} = p_{51},$$

a więc

$$P(T_{50} > 2 \mid T_{50} > 1) = P(T_{51} > 1).$$

Na pierwszy rzut oka wydaje się, że równość taka powinna zachodzić w każdej populacji, gdyż w obydwu przypadkach chodzi o przeżycie od 51 do 52 roku życia. Ale pierwsze z tych prawdopodobieństwo dotyczy populacji 50-latków, a drugie populacji 51-latków. Mogło się tak zdarzyć, że strsze pokolenie 51-latków przeżyło w pierwszym roku życia jakiś kataklizm, który ominął 50-latków, ale zdarzenie to może mieć wpływ na rozkład przyszłego czasu życia.

3. Hipotezy interpolacyjne

Załóżmy, że dany jest rozkład zmiennej losowej K_x dla każdego $x = 0, 1, 2, \dots$, a w szczególności dane są prawdopodobieństwa ${}_np_x$ dla $n, x = 0, 1, 2, \dots$. Hipotezy interpolacyjne umożliwiają wyznaczenie wartości funkcji ${}_tp_x$ dla $t \in [n, n + 1)$, $n = 0, 1, 2, \dots$.

Oznaczmy przez S_x **ułamkowy czas życia**, tzn.

$$S_x = T_x - K_x.$$

Zauważmy, że jeśli $n = 0, 1, 2, \dots$ oraz $u \in [0, 1)$, to

$$P(T_x \leqslant n + u) = P(K_x + S_x \leqslant n + u) = P(S_x \leqslant u \mid K_x = n)P(K_x = n)$$

a więc

$${}_{n+u}p_x = P(S_x \leqslant u \mid K_x = n)({}_np_x - {}_{n+1}p_x).$$

Zatem przyjęcie pewnej hipotezy interpolacyjnej jest równoważne określeniu warunkowego rozkładu S_x przy warunku $K_x = n$.

Definicja 1. Powiemy, że rozkład T_x spełnia **hipotezę jednostajności** (HU), jeżeli funkcja ${}_tp_x$ zmiennej t jest ciągła i liniowa na przedziałach $[n, n+1)$. Zatem

$$ {}_{n+u}p_x = (1-u)_np_x + u \cdot {}_{n+1}p_x, \quad 0 \leqslant u < 1,\ n = 0, 1, 2, \ldots . $$

Zauważmy, że interpolacja jest dokonywana zawsze między kolejnymi latami. Zatem znajomość ${}_3p_{30}$ i ${}_5p_{30}$ nie wystarczy do wyznaczenia ${}_{4.5}p_{30}$. Ale prawdopodobieństwo to można wyznaczyć znając ${}_4p_{30}$ i ${}_5p_{30}$, ze wzoru ${}_{4.5}p_{30} = 0.5({}_4p_{30} + {}_5p_{30})$.

Podstawiając w powyższej definicji $n = 0$ dostajemy

$$ {}_up_x = 1 - u + u\,p_x $$

a więc przy założeniu HU dla $u \in (0, 1)$ mamy

$$ {}_up_x = 1 - uq_x, \quad {}_uq_x = uq_x. $$

Twierdzenie 4. *Niech będzie dany rozkład K_x. Wtedy HU jest równoważna warunkowi*

$$ \mathrm{P}(S_x \leqslant u \mid K_x = n) = u, $$

dla $0 \leqslant u < 1$ *i* $n = 0, 1, 2, \ldots$.

Dowód. Jeżeli zachodzi HU, to

$$ \begin{aligned} \mathrm{P}(K_x = n, S_x \leqslant u) &= \mathrm{P}(n \leqslant T_x \leqslant n+u) = {}_np_x - {}_{n+u}p_x \\ &= {}_np_x - (1-u)_np_x - u \cdot {}_{n+1}p_x \\ &= u({}_np_x - {}_{n+1}p_x) = u\mathrm{P}(K_x = n). \end{aligned} $$

Zatem

$$ \mathrm{P}(S_x \leqslant u \mid K_x = n) = \frac{\mathrm{P}(K_x = n, S_x \leqslant u)}{\mathrm{P}(K_x = n)} = u, $$

co należało pokazać. □

Powyższe twierdzenie mówi, że przy założeniu HU zmienne losowe K_x i S_x są niezależne i S_x ma rozkład jednostajny na przedziale $[0, 1]$ (stąd nazwa hipotezy). W szczególności

$$ \mathring{e}_x = e_x + \frac{1}{2} $$

oraz

$$ \operatorname{Var} T_x = \operatorname{Var} K_x + \frac{1}{12}. $$

Definicja 2. Powiemy, że rozkład T_x spełnia **hipotezę przedziałami stałego natężenia zgonów** (HCFM), jeżeli $\mu_{[x]+t}$ jest funkcją stałą zmiennej t w przedziałach $(n, n+1)$, $n = 0, 1, 2, \ldots$, tzn.

$$\mu_{[x]+n+u} = \mu_{[x]+n}, \quad 0 \leqslant u < 1.$$

Przy założeniu HCFM mamy

$$\mu_{[x]+n+u} = \mu_{[x]+n} = -\log p_{[x]+n}, \quad 0 \leqslant u < 1,\ n = 0, 1, 2, \ldots .$$

Istotnie, korzystając ze wzoru

$${}_tp_x = \exp\left(-\int_0^t \mu_{[x]+s} ds\right)$$

mamy

$${}_np_x = \exp\left(-\sum_{k=0}^{n-1} \mu_{[x]+k}\right).$$

Stąd

$$p_{[x]+n} = \frac{{}_{n+1}p_x}{{}_np_x} = \exp\left(-\mu_{[x]+n}\right),$$

co daje szukaną równość.

Twierdzenie 5. *Niech będzie dany rozkład K_x. Wtedy HCFM jest równoważna każdemu z następujących warunków:*

(a) ${}_{n+u}p_x = {}_np_x \left(p_{[x]+n}\right)^u$;

(b) $\mathrm{P}(S_x \leqslant u \mid K_x = n) = \frac{1-(p_{[x]+n})^u}{q_{[x]+n}}$.

Dowód. Mamy

$$\begin{aligned} {}_{n+u}p_x &= \exp\left(-\int_0^n \mu_{[x]+t} dt - \int_n^{n+u} \mu_{[x]+t} dt\right) \\ &= \exp\left(-\int_0^n \mu_{[x]+t} dt\right) \exp\left(-\int_n^{n+u} \mu_{[x]+t} dt\right) \\ &{}_np_x \exp\left(-\int_n^{n+u} \mu_{[x]+t} dt\right). \end{aligned}$$

Zatem jeżeli zachodzi HCFM, to

$${}_{n+u}p_x = {}_np_x \exp\left(u \log p_{[x]+n}\right) = {}_np_x \left(p_{[x]+n}\right)^u.$$

Korzystając teraz z (a) otrzymujemy

$$\begin{aligned} \mathrm{P}(S_x \leqslant u \mid K_x = n) &= \frac{{}_np_x - {}_{n+u}p_x}{{}_np_x - {}_{n+1}p_x} \\ &= \frac{{}_np_x - {}_np_x \left(p_{[x]+n}\right)^u}{{}_np_x - {}_{n+1}p_x}. \end{aligned}$$

Dzieląc licznik i mianownik przez ${}_np_x$ dostajemy warunek (b).

Jeśli zachodzi (b), to zachodzi HCFM. $\square$

W szczególności przy założeniu HCFM zmienne losowe K_x i S_x nie są niezależne.

Zauważmy jeszcze, że z warunku (a) z $n = 0$ dostajemy dla $0 \leqslant u < 1$

$$ {}_u p_x = (p_x)^u, \quad {}_u q_x = 1 - (p_x)^u. $$

Definicja 3. Powiemy, że rozkład T_x spełnia **hipotezę Balducciego**, jeżeli

$$ {}_{1-u} q_{[x]+n+u} = (1-u) q_{[x]+n}, \quad 0 \leqslant u < 1,\ n = 0, 1, 2, \ldots . $$

Zatem HB mówi, że prawdopodobieństwo tego, że x-latek umrze przed końcem n-trgo roku pod warunkiem, że przeżyje do chwili $n + u$ jest proporcjonalne do pozostałej części roku, tj. $1 - u$.

Twierdzenie 6. *Jeżeli dany jest rozkład K_x, to HB jest równoważna warukowi*

$$ {}_{n+u} p_x = {}_n p_x \frac{p_{[x]+n}}{u + (1-u) p_{[x]+n}}. $$

Dowód. Mamy

$$ {}_{n+1} p_x = {}_{n+u} p_x \cdot {}_{1-u} p_{[x]+n+u}, $$

co implikuje równoważność HB i wzoru z tezy twierdzenia. □

W szczególności kładąc $n = 0$ otrzymujemy, że przy założeniu HB

$$ {}_u p_x = \frac{p_x}{u + (1-u) p_x}, \quad {}_u q_x = \frac{u q_x}{u + (1-u) p_x}. $$

Można pokazać, że również w przypadku HB zmienne losowe S_x i K_x nie są niezależne.

Przykład 10. Zakładając, że zachodzi HJP oraz mając dane $p_{70} = 0.98288$ oraz $p_{71} = 0.98102$, obliczyć prawdopodobieństwo tego, że osoba 70-letnia przeżyje jeszcze 1 rok i 3 miesiące przy założeniu HU, HCFM i HB.

Rozwiązanie. Mamy

$$ \mathrm{P}(T_{70} > 1.25) = {}_{1.25} p_{70} = p_{70}\, {}_{0.25} p_{71}. $$

Przy założeniu HU mamy ${}_u p_x = 1 - u q_x$, a więc

$$ {}_{0.25} p_{71} = 1 - 0.25 \cdot q_{71} = 1 - 0.25(1 - p_{71}) = 0.99525. $$

Zatem $\mathrm{P}(T_{70} > 1.25) = 0.97822$.

Przy założeniu HCFM mamy ${}_u p_x = (p_x)^u$, a więc

$$ {}_{0.25} p_{71} = (p_{71})^{0.25} = 0,99522 $$

oraz $\mathrm{P}(T_{70} > 1.25) = 0.97818$.

Przy założeniu HB mamy

$$ {}_u p_x = \frac{p_x}{u + (1-u) p_x}, $$

a więc

$$ {}_{0.25}p_{71} = \frac{p_{71}}{0.25 + 0.75 \cdot p_{71}} = 0,99519, $$

oraz $\mathrm{P}(T_{70} > 1.25) = 0.97815$.

4. Konstrukcja tablic trwania życia

Niech $K_x = \lfloor T_x \rfloor$ oznacza obcięty przyszły czas trwania życia. **Tablicą trwania życia** dla zmiennej K_x nazywamy zbiór par liczb (k, l_k), $k = 0, 1, 2, \dots$, gdzie

$$ l_k = l_0 \mathrm{P}(K_x \geqslant k), \quad k = 0, 1, 2, \dots, $$

oraz l_0 oznacza początkową liczebność populacji x-latków. Zatem z TTŻ dla K_x można odczytać prawdopodobieństwa

$$ {}_kp_x = \mathrm{P}(K_x \geqslant k) = \frac{l_k}{l_0}, \quad x, k = 0, 1, 2, \dots . $$

Oczywiście w powyższej równości l_0 i l_k zależą od x.

W praktyce podaje się tylko TTŻ dla K_0, a tablice dla pozostałych wartości x wyznacza się korzystając z hipotezy agregacji

$$ \mathrm{P}(K_x \geqslant k) = \mathrm{P}(K_0 \geqslant x + k \mid K_0 \geqslant x). $$

Twierdzenie 7. *Załóżmy, że zachodzi hipoteza agregacji (HA). Niech (k, l_k), $k = 0, 1, 2, \dots$, będzie TTŻ dla zmiennej K_0. Wtedy*

$$ (k, l_{x+k}), \quad k = 0, 1, 2, \dots, $$

jest TTŻ dla zmiennej losowej K_x, $x = 0, 1, 2, \dots$, tzn.

$$ {}_kp_x = \frac{l_{x+k}}{l_x}, $$

o ile $l_x > 0$.

Dowód. Mamy na mocy HA

$$ {}_kp_x = \frac{\mathrm{P}(K_0 \geqslant x + k)}{\mathrm{P}(K_0 \geqslant x)}. $$

Ale

$$ \mathrm{P}(K_0 \geqslant x + k) = \frac{l_{x+k}}{l_0}, \quad \mathrm{P}(K_0 \geqslant x) = \frac{l_x}{l_0}, $$

a więc

$$ {}_kp_x = \frac{l_{x+k}}{l_0} \frac{l_0}{l_x} = \frac{l_{x+k}}{l_x}. $$

$\square$

W praktyce w TTŻ dla K_0 oprócz liczb l_k, $k = 0, 1, 2, \ldots, \omega - 1$, gdzie ω jest wiekeim granicznym w populacji, wypisuje się inne wielkości które można wyrazić za pomocą l_k, np. p_k, q_k, e_k oraz

$$d_k = l_k - l_{k+1},$$

czyli oczekiwaną liczbę osób z początkowej populacji, które umarły w wieku k lat.

Twierdzenie 8. *Niech (k, l_k), $k = 0, 1, 2, \ldots$, będzie TTŻ dla zmiennej losowej K_0 przy założeniu HA. Wtedy*

$$\begin{aligned} q_x &= \frac{d_x}{l_x} = \frac{l_x - l_{x+1}}{l_x}; \\ p_x &= \frac{l_{x+1}}{l_x}; \\ e_x &= \frac{l_{x+1} + l_{x+2} + \ldots}{l_x} \\ &= \frac{l_x + l_{x+1} + l_{x+2} + \ldots}{l_x} - 1. \end{aligned}$$

Dowód. Wzór na p_x wynika z poprzedniego twierdzenia z $k = 1$. Dalej

$$q_x = 1 - p_x = 1 - \frac{l_{x+1}}{l_x} = \frac{l_x - l_{x+1}}{l_x} = \frac{d_x}{l_x}.$$

Ponadto

$$e_x = \frac{1}{{}_xp_0} \sum_{k=x+1}^{\infty} {}_kp_0 = \frac{l_0}{l_x} \sum_{k=x+1}^{\infty} \frac{l_k}{l_0} = \frac{1}{l_x} \left(l_{x+1} + l_{x+2} + \ldots \right). \qquad \square$$

W Polsce tablice trwania życia publikuje corocznie Główny Urząd Statystyczny. Z uwagi na znaczne różnice trwania życia mężczyzn i kobiet, podaje się TTŻ osobno dla każdej płci. W tablicach tych $\omega = 100$ oraz $l_0 = 100000$ i podane w nich są kolejno: x, l_x, q_x, d_x, L_x, T_x oraz e_x.

Wielkości l_x, q_x i d_x oznaczają to samo co powyżej.

Wielkość L_x zwana **ludnością stacjonarną** w wieku x obliczona jest ze wzoru

$$L_x = \frac{l_x + l_{x+1}}{2}.$$

Zauważmy, że

$$\frac{L_x}{l_0} = \frac{p_x + p_{x+1}}{2}$$

a więc L_x jest oczekiwaną liczbą członków populacji, którzy dożyli do chwili $x + 0.5$, przy założeniu HU.

Wielkość T_x zwana **skumulowaną ludnością stacjonarną** w wieku x obliczona jest ze wzoru

$$T_x = \sum_{y \geqslant x} L_y = L_x + L_{x+1} + L_{x+2} + \ldots.$$

Wielkość e_x zwana **przeciętnym dalszym trwaniem życia** obliczona jest ze wzoru

$$e_x = \frac{T_x}{l_x}.$$

Oznaczmy chwilowo przez $\bar{e}_x$ obcięty przyszły czas życia. Wiemy, że

$$\bar{e}_x = \frac{1}{l_x} \sum_{k=x+1}^{\infty} l_k,$$

a więc jak łatwo pokazać

$$e_x = \bar{e}_x + \frac{1}{2} = \mathring{e}_x.$$

Zatem przy założeniu HU welkość e_x występująca w TTŻ GUS jest równa $\mathring{e}_x$, czyli przyszłemu oczekiwanemu czasowi życia.