MODELE MATEMATYCZNE
W UBEZPIECZENIACH

WYKŁAD 3: WYZNACZANIE ROZKŁADU CZASU PRZYSZŁEGO ŻYCIA

1. 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

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

a więc HJP jest równoważna warunkowi

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

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

$${}_tp_x = \frac{s(x+t)}{s(x)}.$$

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

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

□

Hipoteza HJP nie zawsze musi być spełniona. Jeśli bowiem zachodzi np. HA, to na mocy powyższego twierdzenia mamy na przykład

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

a więc

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

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.

2. Przykłady teoretycznych rozkładów 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 $B \geqslant 0$ i $c > 1$ oraz $A \geqslant -B$. W szczególności dla $B = 0$ otrzymujemy rozkład o stałym natężeniu śmiertelności, czyli rozkład wykładniczy (ćw.)
- Rozkład Weibulla (1939), który zakładał, że
$$\mu_t = kt^n, \quad t \geqslant 0,$$
gdzie $k > 0$, $n > 0$.

3. Tablice 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), $k = 0, 1, 2, \ldots$, gdzie
$$l_k = l_0 \mathrm{P}(K_x \geqslant k), \quad k = 0, 1, 2, \ldots,$$

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

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

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 HJP. Mamy wtedy

$$ \begin{aligned} {}_k p_x &= \mathrm{P}(K_x \geqslant k) = \mathrm{P}(K_0 \geqslant x + k \mid K_0 \geqslant x) \\ &= \frac{\mathrm{P}(K_0 \geqslant x + k)}{\mathrm{P}(K_0 \geqslant x)} = \frac{l_{x+k}/l_0}{l_x/l_0} = \frac{l_{x+k}}{l_x}, \end{aligned} $$

a więc

$$ {}_k p_x = \frac{l_{x+k}}{l_x}. $$

Liczby l_k interpretujemy wtedy jako oczekiwaną liczbę członków danej populacji noworodków, którzy dożyją do wieku k lat.

W praktyce w TTŻ dla K_0 oprócz liczb l_k, $k = 0, 1, 2, \dots, \omega - 1$, 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 2. *Niech (k, l_k), $k = 0, 1, 2, \dots$, będzie TTŻ dla zmiennej losowej K_0 przy założeniu HJP. Wtedy*

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

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

$$ e_x = \frac{1}{{}_x p_0} \sum_{k=x+1}^{\infty} {}_k p_0 = \frac{l_0}{l_x} \sum_{k=x+1}^{\infty} \frac{l_k}{l_0} = \frac{1}{l_x} \left(l_{x+1} + l_{x+2} + \dots \right). \qquad \square $$

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

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

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 wielkość e_x występująca w TTŻ GUS jest równa $\mathring{e}_x$, czyli przyszłemu oczekiwanemu czasowi życia.

4. Hipoteza jednostajności

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 $n, x = 0, 1, 2, \dots$. Hipoteza jednostajności HU 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

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

a więc

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

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

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

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 3. *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* $n = 0, 1, 2, \dots$.

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

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

Przykład 1. 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.

Rozwiązanie. Mamy

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

Przy założeniu HU mamy ${}_{u}p_{x} = 1 - uq_{x}$, a więc

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

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