LMD mokslo darbai, 366–371
1998, Vilnius

Lietuvos gyventojų kohortinio išgyvenamumo modeliavimas

R. Mišeikytė (LSIC)

Išgyvenamumo funkcija – tai patikimumo funkcija $S(t)$, vertinanti, kokia nagrinėjamos populiacijos dalis išgyvena iki amžiaus t.

$$S(t) = \exp\Big(-\int_0^t \mu(\tau)d\tau\Big),$$

čia $\mu(t)$ – momentinė rizika numirti, arba tiesiog mirtingumo intensyvumas amžiuje t.

Kohortinė išgyvenamumo funkcija aprašo kohortos išgyvenamumą. Kohorta – tai grupė žmonių, gimusių tuo pačiu laikotarpiu. Galima kalbėti, pavyzdžiui, apie 1930-ų metų kohortą, arba apie kohortą, sudarytą iš 1928–1934 metais gimusių žmonių. Beje, 1928 metais gimę mūsų Respublikos piliečiai yra seniausia Lietuvos kohorta, kurios išmirimą pirmaisiais gyvenimo metais galime sekti netgi mėnesių tikslumu. Šiame darbe ir nagrinėsime kohortas, gimusias po 1927-ųjų.

Duomenys

Sekame kohortas, sudarytas iš Lietuvos gyventojų, gimusių 1928–1935 bei pokario metais. Skaičiavimams naudojami prieškario statistikos metraščių bei pokario rutininės statistikos duomenys (1939–1945 metais duomenų nėra), sugrupuoti amžiaus intervalais, dažniausiai pamečiui. Tai – gimusių skaičius, mirusių skaičius, bei gyvenančių tam tikro amžiaus asmenų skaičiaus įvertis (tikslūs gyventojų skaičiai nežinomi netgi surašymų metais – dėl taip vadinamos akumuliacijos paklaidos). Dar vienas didelių klaidų šaltinis – pokario mechaninis gyventojų judėjimas. Kadangi nepavyko surasti paamžinių migracijos duomenų, tai gyvenančių tam tikro amžiaus asmenų skaičius laikotarpiu iki pirmojo po karo gyventojų surašymo, vykusio 1959 metais galėjo būti atstatytas ne didesniu kaip 25 procentų tikslumu.

Modeliai

Tegul $\mu_i(t)$ žymi i-tojo individo riziką numirti. Pagal senėjimo teoriją [4] ją galima išreikšti suma $\mu_i(t) = \mu_{ji}(t) + \mu_{si}(t)$, čia $\mu_{si}(t)$ – rizika, susijusi su senatve, o $\mu_{ji}(t)$ – rizika numirti dėl nepalankių išorinių veiksnių, sąlygojanti ankstyvąsias mirtis. Kadangi įvairaus amžiaus žmonių mirties priežasčių struktūra labai skiriasi (kūdikiai daugiausia miršta dėl perinatalinių mirties priežasčių bei įgimtų anomalijų, jaunimas ir vidurinioji karta – dėl traumų ir apsinuodijimų, o pagyvenusių žmonių pagrindinė mirties priežastis – kraujo apytakos sistemos ligos), galima būtų tikėtis, kad šios dvi rizikos yra visiškai, ar bent jau iš dalies, nepriklausomos. Jei ši prielaida teisinga, tai ne tik individo, bet ir visos populiacijos išgyvenamumo funkcija išsireiškia kaip dviejų nepriklausomų patikimimo funkcijų sandauga:

$$S(t) = S_j(t) \cdot S_s(t),$$

čia $S_j(t)$ ir $S_s(t)$ yra individualių dalinių išgyvenamumo funkcijų mišiniai: $S_j(t) = E(S_{ji}(t))$ ir $S_s(t) = E(S_{si}(t))$, atitinkamai. Jei, be to, tarsime, kad $\mu_{ji}(t) = \mu_{ji} \equiv const.$ kiekvienam individui ir turi Gamma pasiskirstymą su vidurkiu λ bei dispersija $\rho^2\lambda^2$ populiacijos viduje [2], turėsime:

$$S_j(t) = \left(1 + \sigma^2\lambda t\right)^{-\frac{1}{\sigma^2}} \quad \text{ir} \quad \mu_j(t) = \lambda(1 + \sigma^2\lambda t)^{-1}.$$

Tuomet populiacinė išgyvenamumo funkcija bus:

$$S(t) = \left(1 + \sigma^2\lambda t\right)^{-\frac{1}{\sigma^2}} S_s(t), \sigma^2 > 0, \lambda > 0,$$

čia $S_s(t)$ aproksimuojama viena iš penkių demografijoje naudojamų patikimumo funkcijų šeimų [1], [2], [3]:

$$S_1(t) = \exp\Big[-\Lambda(t)\Big], \qquad S_2(t) = \Big[1 + \Lambda(t)\Big]^{-1},$$

$$S_3(t) = \exp\Big[1 - \exp(-\Lambda(t))\Big], \qquad S_4(t) = \Big[1 + \rho^2\Lambda(t)\Big]^{-\frac{1}{\rho^2}},$$

$$S_5(t) = \exp\Big[-\frac{1-\theta}{\theta\rho^2}((1 + \frac{\rho^2}{1-\theta}\Lambda(t))^\theta - 1)\Big]$$

kur

$$\Lambda_1(t) = \exp\Big[\alpha + \beta\log(t)\Big], \qquad \Lambda_2(t) = \Lambda_1\Big[\frac{t}{H-t}\Big],$$

$$\Lambda_3(t) = \exp\Big[\alpha + \beta\frac{t^k - t^{-k}}{2k}\Big], \qquad \Lambda_7(t) = \Lambda_3\Big[\frac{t}{H-t}\Big],$$

$$\Lambda_4(t) = \exp\Big[\alpha + \beta\log(\log(1 + t^k))\Big], \qquad \Lambda_6(t) = \Lambda_4\Big[\frac{t}{H-t}\Big],$$

$$\Lambda_5(t) = \exp\Big[\alpha + \beta\log(e^{kt} - 1)\Big], \qquad \Lambda_8(t) = \Lambda_5\Big[\frac{t}{H-t}\Big],$$

$$\Lambda_9(t) = -Q\log\Big[1 - \frac{e^{\gamma t} - 1}{e^{\gamma H} - 1}\Big].$$

$0 < \theta < 1$, $\rho^2 > 0$, α, $\beta > 0$, $H > t$, $k > 0$, $Q > 0$ ir $\gamma > 0$ yra, atitinkamai, funkcijų šeimų bei transformacijų parametrai.

Modelį, gautą įstačius į funkcijų šeimą S_k transformaciją Λ_l, žymėsime S_{kl}.

Metodai

Pažymėkime m_i – mirusių i-tajame intervale skaičių, N_i – tikslų gyvų i-tojo intervalo pradžioje asmenų skaičių ir q_i – tikimybę numirti tame intervale. Tada laukiamas mirusiųjų skaičius i-tajame intervale bus

$$\lambda_i := E(m_i) = N_i * q_i.$$

Jei aproksimuosime m_i pasiskirstymą Puasono skirstiniu, tai tikėtinumo santykio logaritmas $\log(L_i)$ bus

$$\log(L_i) := m_i \log(\lambda_i/m_i) - \lambda_i + m_i,$$

ir

$$-2\log(L_i) \backsim \chi^2(1)$$

Deja, vietoje N_i stebime tiktai jo įvertį n_i. Pažymėję

$$\delta_i^2 := E(\frac{N_i - n_i}{n_i})^2 \quad \text{ir} \quad lauk_i := n_i * q_i$$

gausime, kad

$$\begin{aligned} E(-2log(L_i')) &= 1 + E(2(\lambda_i \log(\lambda_i/lauk_i) + lauk_i - \lambda_i)) \\ &\approx 1 + \lambda_i\delta_i^2 \approx 1 + m_i\delta_i^2 \sim 1 + m_i\Delta_i^2, \end{aligned}$$

čia $\log(L_i') := m_i \log(lauk_i/m_i) - m_i + lauk_i$, o Δ_i^2 yra koks nors turimas (galbūt euristiškai išmąstytas) δ_i^2 įvertis.

Pažymėkime

$$\rho_i := 2\frac{m_i \log(m_i/lauk_i) + lauk_i - m_i}{1 + m_i\Delta_i^2}$$

ir Gr – amžiaus grupių, kuriose turime kohortos stebėjimus, skaičių. Aišku, kad skirtingoms kohortoms šie skaičiai skirsis – vėliau gimusios kohortos stebėjimų turės mažiau.

Tuomet funkcija

$$Q' := \sum_{i=1}^{Gr} \begin{cases} \rho_i, & \text{kai } \rho_i < 8, \\ 6 + \sqrt{2\rho_i - 12}, & \text{priešingu atveju} \end{cases} \tag{1}$$

tam tikru būdu apibrėžia atstumą nuo tikrų duomenų iki modelio.

Ši funkcija buvo minimizuota vyrams ir moterims bei kiekvienai kohortai atskirai. Buvo vertinamos parametrų reikšmės 1928–1935 ir 1946–1966 metais gimusioms kohortoms (iš viso 29-ioms kohortoms). Modelių palyginimui pasirinktas kokybės funkcionalas:

$$\Xi := \frac{1}{58}\sum_{lyt=1}^{2}\sum_{i=1}^{29}\frac{1}{Gr_i - p}Q'_{i,lyt}$$

čia $Q'_{i,lyt}$ yra i-tosios kohortos lyt lyties minimizuota funkcija (1), Gr_i – stebėtų metų skaičius i-tajai kohortai ir p – modelio parametrų skaičius.

Rezultatai

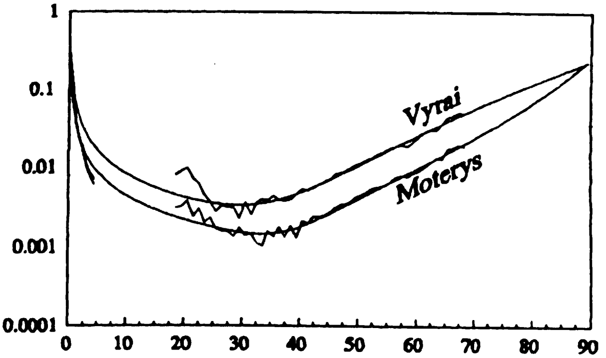
Funkcionalo Ξ įverčiai nagrinėjamoms funkcijų šeimoms bei transformacijoms pateikti šioje lentelėje:

Ξ_{kl}	S_1	S_2	S_3	S_4	S_5	$\Xi_{\cdot l}$
Λ_1	1.30	1.37	1.28	1.33	1.19	1.30
Λ_2	1.28	1.29	1.30	1.29	1.10	1.25
Λ_3	1.29	1.40	1.30	1.37	1.07	1.29
Λ_4	1.34	1.41	1.30	1.37	1.28	1.34
Λ_5	1.46	1.36	1.51	1.51	1.14	1.40
Λ_6	1.36	1.38	1.35	1.36	1.15	1.32
Λ_7	1.16	1.16	1.24	1.19	1.15	1.18
Λ_8	1.31	1.32	1.33	1.35	1.11	1.28
Λ_9	1.48	1.47	1.50	1.36	1.16	1.40
$\Xi_{k\cdot}$	1.33	1.35	1.35	1.35	1.15	1.31

čia k ir l, atitinkamai, funkcijų šeimos bei transformacijos numeriai, o lentelės paskutiniame stulpelyje bei eilutėje pateikti, atitinkamai, eilučių bei stulpelių vidurkiai – kad būtų galima greitai palyginti ne tik konkrečius modelius, bet ir transformacijas Λ_l arba funkcijų šeimas S_k.

Kaip matyti iš grafiko, labiausiai į akis krentantis modelio ir duomenų neatitikimas stebimas 5–30 metų amžiaus grupėje. Tai kaip tik atitinka laikotarpį iki 59-ųjų metų surašymo ir todėl gali būti aiškinama gyventojų skaičiaus netikslumais. Nežiūrint to, panašūs, nors ir mažiau ryškūs, nukrypimai stebimi ir kitose kohortose. Kadangi didesnę šios amžiaus grupės mirtingumo dalį sudaro išorinės mirties priežastys, tai natūralu tikėtis, kad nemaža dalimi šiuos nukrypimus įtakoja nežinomi išoriniai veiksniai, galintys priklausyti ir nuo amžiaus, ir nuo medicinos išsivystymo lygio, ir nuo politinės situacijos.

Palyginimui pateikti 1988–1996 metų sintetinės kohortos modeliavimo rezultatai, kai į galimus gyventojų skaičiaus netikslumus neatsižvelgiama (laikotarpinis išgyvenamumas,



Pav. 1928-ųjų metų gimimo kohortos mirtingumas (logaritminė skalė). Išgyvenamumo funkcija aproksimuota modeliu $S(t) = S_j(t)S_{17}(t)$.

arba sintetinės kohortos išgyvenamumas, – tai išgyvenamumas įsivaizduojamos pseudopopuliacijos, kurios mirtingumas visose amžiaus grupėse sutampa su tiriamo laikotarpio atitinkamų amžiaus grupių mirtingumu). Šiuo atveju maksimizuojamas maksimalaus tikėtinumo santykio logaritmas vyrams ir moterims atskirai:

$$\log(L) := \sum_{j=0}^{99} m_j \log\left(\frac{lauk_j}{mj}\right) - lauk_j + m_j.$$

čia m_j ir $lauk_j$ žymi tą patį, kaip ir anksčiau, tik laikotarpiniams duomenims.

Antroje lentelėje pateiktos funkcionalo

$$\Xi' := -\frac{2}{100-p}\sum_{lyt=1}^{2} log(L_{lyt})$$

reikšmės (kai modelis parinktas tinkamai ir gyventojų skaičiai amžiaus grupėse tikslūs $E(\Xi') = 1$).

Ξ'_{kl}	S_1	S_2	S_3	S_4	S_5	$\Xi'_{\cdot l}$
Λ_1	24.59	84.95	5.94	24.85	25.11	33.09
Λ_2	2.76	4.19	4.05	2.11	1.93	3.01
Λ_3	3.30	25.80	3.50	3.36	2.01	7.59
Λ_4	26.57	85.18	6.00	27.44	25.42	34.12
Λ_5	5.51	56.08	13.56	5.62	4.40	17.03
Λ_6	2.63	5.12	2.69	2.09	1.67	2.84
Λ_7	2.70	2.32	4.09	2.06	1.93	2.62
Λ_8	2.54	2.55	3.91	2.01	1.83	2.56
Λ_9	5.68	44.09	16.97	4.27	3.11	14.82
$\Xi'_{k\cdot}$	8.47	34.47	6.75	8.20	7.49	13.08

Gauti rezultatai leidžia konstantuoti, kad, vykdant tolesnę kohortinių duomenų analizę, verta apsiriboti funkcijomis S_{52}, S_{53}, S_{56}, S_{57} bei S_{58}. Funkcijos S_{51}, S_{55}, S_{59}, S_{17} ir S_{27} nors ir duoda neblogus rezultatus aproksimuojant turimus kohortų duomenis (apimančius ne vyresnes kaip 66 metų amžiaus grupes), vargu ar gali patenkinamai aproksimuoti vyresnio amžiaus žmonių mitringumą (nes tokiu atveju laikotarpinių duomenų, apimančių visą 0–100 metų amžiaus intervalą aproksimacija taip pat būtų tikslesnė).

LITERATŪRA

[1] D.J. Slimen and P.A. Lasenbruch, Survival distributions arising from two families and generated by transformations, *Commun. Statist.*: *Theory and Meth.*, **13** (10) (1984), 1179–1201.

[2] P. Hougaard, Fraitly Models for Survival Data, *Lifetime Data Analysis*, (1) (1995), 255–273.

[3] V. K. Koltover, Z. S. Andrianova, and A. N. Ivanova, Modeling of survival and mortality curves of human population based on the theory of reliability, *Nauka, Ser. Biol.*, (1) (1993 (rusų kalba)), 121–129.

[4] D. Zelterman, A Statistical distribution with an Unbounded Hazard Function and its Application to a Theory from Demography, *Biometrics*, (48) (1992), 807–818.

Modeling of Lithuanian cohort survivorship

R. Mišeikytė

Survival distributions, arising from six parametrical families were fitted to the mortality experience of Lithuanian cohorts, born in 1928–1935 and 1946–1966 years.