LMD mokslo darbai, III tomas, 367–369

Daugiamatis logistinis mirčių prognozavimo modelis

Audronė JAKAITIENĖ (VDU)
el. paštas: iiauja@vdu.lt

Sveikatos priežiūros rezultato, t.y. poveikio gyventojų sveikatai tiesiogiai išmatuoti neįmanoma, tačiau mirtingumo dėsningumai gali padėti nustatyti sveikatos priežiūros efektyvumo trūkumus populiaciniu lygiu ir numatyti būdus, kaip juos pašalinti. Dėsningumai turi būti nustatyti remiantis populiacijos individų pasiskirstymu pagal objektų požymius. Remiantis šiais dėsningumais galima kiekybiškai įvertinti pasiektus sveikatos priežiūros tikslus, jų trūkumus ir numatyti galimus pakeitimus. Taigi, galutiniai Lietuvos gyventojų sveikatos būklės vertinimai remtųsi mirtingumo statistika.

Vienas iš būdų mirtingumui mažinti yra *išvengiamo* mirtingumo mažinimas. Išvengiamos mirtys – tai mirtys nuo ligų, sąlygotų nesamų ir neveiksmingų profilaktinių priemonių ir/ar nesavalaikės diagnostikos, ir/ar neadekvatyvaus gydymo. Tad, esant pakankamai geram medicinos mokslo lygiui, daugelio mirčių būtų galima išvengti ar jų skaičių ženkliai sumažinti, jeigu būtų imamasi žinomų profilaktinių priemonių arba laiku taikomas šiuolaikinis adekvatus gydymas.

Šiame darbe mirčių priežastims prognozuoti taikomas regresinis multinominis logit modelis (logit-modelis). Šis modelis leidžia, atsižvelgiant į rizikos faktorių pasiskirstymą, prognozuoti vidutinį identifikuojamų mirčių skaičių tiriamoje populiacijoje.

Tegu diskretus atsitiktinis dydis Y_i žymi i-tojo individo mirties priežastį (priežastys arba kategorijos sunumeruotos nuo 1 iki m). Tikimybė, kad i-tasis individas mirs nuo j-tosios priežasties nusakoma taip:

$$\pi_{ij} = \mathbf{P}(Y_i = j) = \frac{e^{x_i' \cdot \gamma_j}}{\sum\limits_{l=1}^{m} e^{x_i' \cdot \gamma_l}}, i = 1, \ldots, n, j = 1, \ldots, m,$$

kur x_i' yra k-matis regresorių vektorius i-tajam stebėjimui ir γ_j yra j-tos kategorijos regresijos parametrų vektorius. Reikalausime, kad $\sum\limits_{j=1}^{m} \pi_{ij} = 1$ ir $\sum\limits_{j=1}^{m} \gamma_j = 0$.

Tikėtinumo lygčiai sudaryti įveskime fiktyvius kintamuosius, $W_{i1}, \ldots, W_{im}$taip, kad $W_{ij} = 1$, jeigu $Y_i = j$, ir $W_{ij} = 0$, jeigu $Y_i \neq j$. Tuomet stebėjimo y_i tikėtinumo funkcija yra

$$p(y_i) = \prod_{j=1}^{m} \pi_{ij}^{W_{ij}}.$$

Jeigu stebėjimai yra nepriklausomi, tai jų bendra tikėtinumo funkcija bus:

$$p(y_1, ..., y_n) = \prod_{i=1}^{n} \prod_{j=1}^{m} \pi_{ij}^{W_{ij}}.$$

Tada modelio tikėtinumo funkcija imčiai $y_1, ..., y_n$ su regresorių matrica $X = (x_1', ..., x_n')$turės pavidalą:

$$L(\gamma_1, ..., \gamma_m) = p(y_1, ..., y_n \mid X) = \prod_{i=1}^{n} \prod_{j=1}^{m} \left(\frac{e^{x_i' \cdot \gamma_j}}{\sum_{l=1}^{m} e^{x_i' \cdot \gamma_l}} \right)^{W_{ij}},$$

o jos logoritmas bus lygus

$$\begin{aligned} \log L(\gamma_1, ..., \gamma_m) &= \sum_{i=1}^{n} \sum_{j=1}^{m} W_{ij} \cdot \left[x_i' \cdot \gamma_j - \log \left(\sum_{l=1}^{m} e^{x_i' \cdot \gamma_l} \right) \right] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} W_{ij} \cdot x_i' \cdot \gamma_j - \sum_{i=1}^{n} \log \left(\sum_{l=1}^{m} e^{x_i' \cdot \gamma_l} \right). \end{aligned}$$

Nežinomų parametrų reikšmės išskaičiuojamos naudojantis iteraciniu Niutono-Rafsono metodu:

$$\hat{c}_{l+1} = \hat{c}_l + \left(-\frac{\partial^2}{\partial^2 \hat{c}_l} \log L(\hat{c}_l) \right)^{-1} \frac{\partial}{\partial \hat{c}_l} \log L(\hat{c}_l),$$

kur $\hat{c}_l$ yra nežinomų parametrų $[k \times (m-1)]$ matricos l-tasis (Niutono-Rafsono arba maksimalaus tikėtinumo) artinys. Pradinės parametrų reikšmės buvo parenkamos, panaudojant daugiamates gardeles.

Simuliacinio modeliavimo būdu buvo sudaryta 1000 atvejų populiacija su trimis mirčių priežastimis (koronarinė mirtis, mirtis nuo vėžio, kita mirtis) ir keturiais požymiais (sistolinis kraujo spaudimas, cholesterinas, gliukozė, kūno masės indeksas (KMI)) su prielaida, kad požymiai klasėse turi normaliuosius skirstinius su skirtingais vidurkiais (medicinos praktikoje nusistovėję dydžiai) ir vienoda dispersija.

Gautus rezultatus, pritaikius modelį šiems duomenims, pateikiame lentelėje.

1 lentelė.
Stebėtų ir apskaičiuotų, panaudojus tiesiniu logistiniu modeliu suskaičiuotus koeficientus, mirčių kiekių palyginimas atitinkamai kiekvienai mirties priežasčiai.

	Koronarinių mirčių sk.	Mirčių nuo vėžio sk.	Kitų mirčių skaičius	Gyvas
Mirčių tikimybių, apskaičiuotų panaudojus tiesinio logistinio modelio suskaičiuotus koeficientus, suma atitinkamai kiekvienai mirties priežasčiai	254	251	253	242
Stebėti mirčių kiekiai atitinkamai kiekvienai mirties priežasčiai	252	241	244	263

Modelis leidžia labai tiksliai prognozuoti laukiamus mirčių skaičius. Skirtumas tarp mirčių tikimybių, apskaičiuotų panaudojus tiesinio logistinio modelio suskaičiuotus koeficientus, sumų ir stebėtų mirčių kiekių atitinkamai kiekvienai mirties priežasčiai yra nedidelis. Remiantis tokiais gautais rezultatais, šį modelį būtų tikslinga toliau nagrinėti dirbant su realiomis klinikinėmis studijomis. Tokiu būdu jis būtų taikomas nustatant atskirų individų mirties tikimybes ir šių tikimybių dydžio priklausomybę nuo rizikos faktorių reikšmių. Ši informacija turėtų būti naudojama bendrosios praktikos gydytojų, kad būtų taikomas kuo adekvatesnis gydymas, ir gyvybės draudimo agentų, kad turėtų pilnesnę informaciją apie draudžiamą klientą.

Regresinio tipo mirčių prognozavimo modeliai leidžia išskirti padidintos rizikos grupes ir taikyti adekvatų gydymą. Tai mažintų *išvengiamą* mirtingumą ir gerintų Lietuvos gyventojų sveikatos būklę.

Literatūra

[1] John Fox, *Linear Statistical Related Methods with Applications to Social Research*, USA (1984).
[2] Regina C. Elandt-Johson, Norman L. Johson, *Survival Models and Data Analysis*, John Wiley & Sons (1979).
[3] Rupert G. Miller, Jr., *Survival Analysis*, John Wiley & Sons (1974).

Multinomial logit death forecasting model

A. Jakaitienė

The multinomial regression logit model is analyzed. The algorithms and software are made for this model in order to get estimation of parameters. Calculations are made using generated population of 1000 cases.