Liet. matem. rink., T. 40, spec. nr., 389–391


# Konkuruojančių rizikų fiksuoto laiko modelis

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

Konkuruojančių rizikų teorija yra susijusi su tam tikros rizikos įvertinimu kitų rizikų atžvilgiu. Konkuruojančių rizikų teorijoje naudojami duomenys, kuriuose tarp stebėjimų yra užfiksuotos mirties priežastys ir mirčių laikas. Visi populiacijos nariai yra veikiami tų pačių $m$, kai $m \geqslant 2$, rizikos faktorių. Tikslas yra išskirti konkrečios rizikos ar rizikų įtaką turimoje populiacijoje. Konkuruojančios rizikos buvo pradėtos naudoti jau 1760 metais. Tuo metu, D. Bernoulli, motyvuodamas skiepų reikalingumą nuo kiaulytės susirgimo, sukūrė konkuruojančių rizikų teorijos pagrindus (žiūr. [1, 2]).

Šiandien mes nežinome, kokių konkrečių rizikos faktorių eliminavimas padėtų Lietuvoje išvengti ar sumažinti mirčių skaičių nuo vėžio, širdies ar kitų šiandien aktualių ligų. Tuo šis uždavinys tampa ypatingai aktualus.

Šiame darbe konkuruojančioms rizikoms – mirčių priežastims – prognozuoti yra taikomas regresinis multinominis logit modelis. Šis modelis leidžia, atsižvelgiant į rizikos faktorių pasiskirstymą, prognozuoti vidutinį identifikuojamų mirčių skaičių tiriamoje populiacijoje . Ankstesniame darbe [6] regresinis multinominis logit modelis ir jo vertinimui sudaryta programinė įranga buvo tirti simuliacinio modeliavimo būdu. Šiame darbe konkuruojančių rizikų prognozavimui naudojami realūs Kardiologinio instituto bazėje ilgalaikiais išgyvenamumo stebėjimais surinkti duomenys apie Kauno m. vyrų sveikatą.

Tegu diskretus atsitiktinis dydis $Y_i$ žymi $i$-tojo individo mirties priežastį. Tikimybė, kad $i$-tasis individas mirs nuo $j$-tosios priežasties nusakoma taip:

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

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

Modelis (1) yra vadinamas daugiamačiu logit modeliu (žiūr. [2, 3, 4, 5]).

Tikėtinumo lygčiai sudaryti įveskime fiktyvius kintamuosius $d_{i1}, \dots, d_{im}$ taip, kad $d_{ij} = 1$, jeigu $Y_i = j$, ir $d_{ij} = 0$, jeigu $Y_i \neq j$. Tuomet

$$p(y_i) = \prod_{j=1}^{m} \left(\mathbf{P}(Y_i = j)\right)^{d_{ij}}.$$

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

$$p(y_1,\dots,y_n)=\prod_{i=1}^{n}\prod_{j=1}^{m}\left(\mathbf{P}(Y_i=j)\right)^{d_{ij}}.$$

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

$$p\left(y_1,\dots,y_n|X\right)=\prod_{i=1}^{n}\prod_{j=1}^{m}\left(\frac{e^{x_i'\cdot\gamma_j}}{\sum\limits_{l=1}^{m}e^{x_i'\cdot\gamma_l}}\right)^{d_{ij}},$$

o jos logaritmas bus lygus

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

Multinominio logit modelio parametrų didžiausio tikėtinumo įverčius apskaičiuosime naudodami Niutono metodą, kuriuo paprastai lengvai randamas sprendinys, nebent duomenys būtų blogai sąlygoti. Pradines parametrų reikšmes parinksime, panaudojant daugiamates gardeles.

Šis modelis buvo pritaikytas Kardiologinio instituto bazėje 1972–1977 m. surinktiems duomenims. Tuomet buvo ištirta 2452 atsitiktinai atrinktų Kauno m. 45–60 amžiaus vyrų sveikata. Ši populiacija buvo stebėta 20m. Šiam darbui iš turimų duomenyse 18 požymių buvo paimti trys požymiai: sistolinis kraujo spaudimas ir cholesterino bei gliukozės kiekis kraujyje, o taip pat išskirtos 3 mirties priežasčių grupės (konkuruojančios rizikos): koronarinė mirtis, mirtis nuo vėžio, mirtys nuo kitų ligų.

Gautus rezultatus, pritaikius modelį šiems duomenims, pateikiame 1 lentelėje.

1 lentelė. Stebėtų ir prognozuotų mirčių skaičių palyginimas

| | Koronarinių mirčių sk. | Mirčių nuo vėžio sk. | Kitų mirčių sk. | Gyvas |
|---|---|---|---|---|
| Prognozuotas N | 513 | 247 | 209 | 1483 |
| Stebėtas N | 457 | 257 | 222 | 1516 |

Modelis leidžia gana tiksliai prognozuoti prognozuojamus mirčių skaičius populiacijoje. Mirčių skaičius populiacijoje nuo $j$-tos priežasties buvo prognozuojamas dydžio $\sum_{i=1}^{n} \mathbf{P}(Y_i = j)$ įverčiu (t.y. suma įvertintų tikimybių nuo konkrečios priežasties (rizikos faktoriaus)). Skirtumas tarp mirčių tikimybių sumų, apskaičiuotų panaudojus tiesinio logistinio modelio suskaičiuotus koeficientus, ir stebėtų mirčių kiekių atitinkamai kiekvienai mirties priežasčiai yra nedidelis. Šį modelį būtų tikslinga patikslinti didinant regresorių kiekį. Sekantis žingsnis būtų rasti tokį regresorių rinkinį, kuris padėtų išvengti ar ženkliai sumažintų atitinkamų mirčių skaičių. Patikslinta informacija galėtų būti naudojama bendrosios praktikos gydytojų, nes prognozė padėtų parinkti adekvatesnį gydymą.

## Literatūra

[1] H.A. David, M.L. Moeshberger, *The Theory of Competing Risks*, Charles Griffin & Company ltd. (1978).
[2] P. Armitage, Th. Colton, *Encyclopedia of Biostatistics*, John Wiley and Sons (1998).
[3] H.W. Greene, *Econometric Analysis*, Prentice Hall (2000).
[4] R.C. Elandt-Johson, N.L. Johson, *Survival Models and Data Analysis*, John Wiley & Sons (1979).
[5] R.G. Miller, Jr., *Survival Analysis*, John Wiley and Sons (1974).
[6] A. Jakaitienė, Daugiamatis logistinis mirčių prognozavimo modelis, *LMD mokslo darbai*, III tomas, 367–369 (1999).

# Fixed time competing risk model

A. Jakaitienė

The competing risks under 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 data of Cardiology Institute about health of 45–60 years old men, which were collected from 1972 till 1977.