УДК 681.3

Катерина Рисцова
Алла Рисцова

МАТЕМАТИЧНА МОДЕЛЬ СТРАХУВАННЯ НА ОСНОВІ УЗАГАЛЬНЕНОЇ ЛІНІЙНОЇ РЕГРЕСІЇ З УРАХУВАННЯМ РИЗИКОВИХ КОМПОНЕНТІВ

У статті наводиться побудова математичної моделі для задачі нарахування страхових премій, розглянуті та обчислені три моделі для страхування з урахуванням ризиків.

В статье приведено построение математической модели для задачи начисления страховых премий, рассмотрены и рассчитаны три модели для страхований с учетом рисков.

Generalized linear regression models (GLM) are considered for the estimation of insurance premium rates. Results are obtained for three exponential distributions of data with the same risk criteria.

Ключові слова: математична модель, лінійна регресія, функція щільності, система нелінійних рівнянь, задача страхування, матриця, ваговий коефіцієнт.

Класична лінійна регресійна модель подається рівнянням $Y = X\beta + \varepsilon$, де *Y*– вектор значень залежної змінної, *X* – матриця значень незалежних змінних, β – вектор регресійних коефіцієнтів, що слід знайти, ε – випадковий вектор відхилень. У цьому випадку рішення проводиться за методом найменших квадратів (LSS).

При використанні лінійної моделі виявилися її недоліки, що звужують коло її застосування. Так, припускається, що усі компоненти змінної *Y* розподілені за нормальним законом з однаковою дисперсією та вектор математичних очікувань *E(Y)* лінійно подається через стовпці незалежних змінних *E(Y)* = $X\beta$. Такі припущення часто не виконуються на практиці і роблять таку модель неадекватною до процесу.

Узагальнена лінійна модель містить у собі комплекс моделей, де класична модель виконується як один з випадків [1]. Узагальнена модель базується на двох положеннях: усі компоненти вектора *Y* незалежні і кожна з них задається імовірним розподіленням; взаємодія між випадковою компонентою *Y* та систематичною компонентою $\eta = X\beta$ здійснюється за допомогою спеціальної функції g, що вважається диференційованою, монотонною та взаємно однозначною.

Група експоненціальних розподілень задається як двопараметрична група функцій щільності: $f(x,\theta,\varphi)=\exp\left\{\frac{x\theta-b(\theta)}{a(\varphi)}+c(x,\varphi)\right\}$, де функції $a(\varphi)$, $b(\theta)$, $c(x,\varphi)$ задаються заздалегідь, параметр θ (канонічний) пов'язаний з математичним очікуванням, а параметр φ пов'язаний з дисперсією.

Для розв'язання узагальненої лінійної регресійної моделі:

$$E(Y)=g^{-1}(X\beta)\ , \tag{1}$$

застосовуємо метод максимальної правдоподібності. Якщо компоненти вектора Y розподілені за одним з експоненціальних законів, то логарифм його функції правдоподібності буде мати вигляд:

$$Lg(Y,\theta)=\sum_{i=1}^{m}\left\{\frac{y_i\theta_i-b(\theta_i)}{a(\varphi)}+c(y_i,\varphi)\right\}. \tag{2}$$

Тепер необхідно виразити параметри θ_i, $1 \le i \le m$, через регресійні коефіцієнти β_j, $1 \le j \le n$,

$$\mu_i = g^{-1}(\beta_1 x_{i1}+\ldots+\beta_n x_{in})\ ,\quad 1\le i\le m\ . \tag{3}$$

Після обчислення часткових похідних отримуємо систему нелінійних рівнянь:

$$\frac{\partial Lg\,(Y,\beta)}{\partial\beta_j}=\sum_{i=1}^{m}\frac{\omega_i(y_i-\mu_i)x_{ij}}{V(\mu_i)g'(\mu_i)}=0\quad ,\quad 1\le j\le n. \tag{4}$$

Для конкретних розподілень та функцій зв'язку ця система значно спрощується: для логарифмічної функції зв'язку $g(x) = \log(x)$ в разі розподілення Пуассона маємо: $V(\mu_i)g'(\mu_i)=1$, а в разі гамма-розподілення маємо $V(\mu_i)g'(\mu_i)=\mu_i$. Таким чином, отримані спрощені системи рівнянь:

$$\sum_{i=1}^{m}\omega_i(y_i-\mu_i)x_{ij}=0\,,\ \ \sum_{i=1}^{m}\omega_i(\frac{y_i}{\mu_i}-1)x_{ij}=0\,,\ 1\le j\le n \tag{5}$$

Підставивши замість μ_i їх вираз (3), отримаємо систему нелінійних рівнянь відносно коефіцієнтів регресії. Рішення цієї системи дають оцінки для коефіцієнтів регресії та параметрів розподілення.

Наведена модель застосовувалась для рішення задачі страхування з урахуванням компонент ризику.

Вектор значень Y, структурна матриця $X=[x_{ij}]$, вагові коефіцієнти ω, обчислюються за вхідними даними та прикладними міркуваннями. Типовим для функції зв'язку в страхуванні є $g(x) = \log(x)$. В цьому випадку:

$$\mu_i=\exp(\beta_1x_{i1}+\ldots+\beta_nx_{in})=\exp(\beta_1x_{i1})\cdot\ldots\cdot\exp(\beta_nx_{in})\ ,\ 1\le i\le m\ . \tag{6}$$

Перший член в добутку відповідає базовій вартості страхування, він сталий. Решта членів являють собою надбавки, пов'язані з різними видами ризику у даній страховій групі.

Для порівняння, розраховувалося і логнормальне розподілення, розв'язане за методом найменших квадратів.

Таким чином, наведено три моделі для розрахунку страхових премій, загальні параметри розрахунків наведені у табл. 1.

Таблиця 1

Модель	**LSM**	**GLM1**	**GLM2**
Розподілення	Логнормальне	Пуассона	Гамма
Функція варіації	1	x	x^2
Функція зв'язку	1	Log(x)	Log(x)
Ваги	1	Розмір вибірки	Розмір вибірки

За початкові дані обиралась вибірка з 2300 записів по автомобільним страховим виплатам. Дані розбиті на 20 груп за значенням 5 факторів: тип авто, регіон страхування, колір авто та інші. Дані підготовлені у форматі Excel, після чого таблиця транспортувалась у програму Maple для обчислення регресійних коефіцієнтів.

Отримані результати наведені в табл. 2.

Таблиця 2

LSM			GLM1				GLM2	
Cr	E(Y)	log(E(Y))	Cr	E(Y)	log(E(Y))	Cr	E(Y)	log(E(Y))
9,108985	9036,117	9,108985	8,898476	7320,811	8,898476	8,892803	7279,395	8,892803
1,196915	7471,055	8,918792	1,55547	6317,18	8,751028	1,556463	6311,605	8,750145
0,376163	7732,364	8,95317	0,762007	6847,078	8,831577	0,766471	6838,287	8,830293
-0,11049	6393,114	8,762977	-0,15533	5908,392	8,684129	-0,15033	5929,142	8,687635
-0,15581	8090,863	8,998491	-0,0669	6267,574	8,743145	-0,06251	6263,371	8,742474
-0,19019	6689,52	8,808297	-0,14745	5408,334	8,595696	-0,14266	5430,66	8,599816
	13162,78	9,485148		15685,37	9,660484		15666,4	9,659274
	10882,98	9,294955		13535,02	9,513035		13583,57	9,516616
	11263,62	9,329333		14670,36	9,593584		14717,07	9,596763
	9312,754	9,13914		12659,16	9,446136		12760,45	9,454105
	11785,84	9,374654		13428,73	9,505152		13479,76	9,508944
	9744,525	9,184461		11587,75	9,357704		11687,63	9,366287
	29908,56	10,3059		34680,98	10,45395		34519,03	10,44927
	24728,38	10,11571		29926,46	10,3065		29929,75	10,30661
	25593,28	10,15009		32436,75	10,38705		32427,29	10,38676
	21160,51	9,959892		27989,9	10,2396		28116,1	10,2441
	26779,87	10,19541		29691,46	10,29861		29701,02	10,29894
	22141,58	10,00521		25620,97	10,15117		25752,29	10,15628
	37281,35	10,52625		69498,09	11,14905		69788,62	11,15323
	30824,2	10,33606		59970,39	11,00161		60510,28	11,01057

ЛІТЕРАТУРА

1. Anderson D. A Practitioner's Guide to Generalized Linear Models // CAS Discussion Paper Program. – N.Y.: Wiley, 2004. – P. 1 –115.