

СТАТИСТИЧЕСКИЙ АНАЛИЗ ЭФФЕКТИВНОСТИ ПРОГРАММ ОТКАЗА ОТ КУРЕНИЯ

Дуплякин В.М.

*Самарский государственный аэрокосмический университет,
г. Самара, Россия*

**Поваляева Л.В., Чугунова И.В., Григошкин С.С., Самойлова Н.В.,
Дуплякина В.Т., Гостева С.Ю.**

*Государственное бюджетное учреждение здравоохранения Самарской
области "Самарская городская больница № 4"), г. Самара, Россия*

Рассматривается эффективность назначения программ отказа от курения на основе статистического анализа выборочных данных кабинета отказа от курения городской больницы СГБ-4 (г. Самара) за период 2012-2013 г. (объём выборки $n = 218$ - достаточно представительная выборка, учитывая, что кабинет работает с середины 2012 года и всего зафиксировано 489 обращений, включая повторные обращения).

При назначении программы отказа от курения в соответствии с утверждёнными методиками проводится всестороннее тестирование пациентов, которое сводится к регистрации совокупности следующих факторов:

Z1 – пол: 1 – мужской, 2 – женский; Z2 – возраст (годы),

Z3 – индекс курильщика (число сигарет, выкуриваемых за год);

Z4 – степень никотиновой зависимости (баллы);

Z5 – степень мотивации бросить курить (баллы);

Z6 – мотивация к курению (баллы);

Z7 – назначаемая программа: 1 – снижение интенсивности курения и повышение мотивации бросить курить, 2 – короткая программа лечения от никотиновой зависимости, 3 – длительная программа;

Y – полученный результат: 1 – отказ от программы, 2 – снижение интенсивности курения, 3 – отказ от курения.

Полнофакторная статистическая модель выбора программы отказа от курения может выглядеть следующим образом

$$Y = b_0 + b_1Z_1 + b_2Z_2 + b_3Z_3 + b_4Z_4 + b_5Z_5 + b_6Z_6 + b_7Z_7. \quad (1)$$

Статистика выборочных данных представлена в таблицах 1 и 2.

Таблица 1 – Структура посетителей кабинета отказа от курения

ПОЛ, %		Категория возраста, лет, %			
МУЖ	ЖЕН	<30	[30-50)	[50-60]	>60
68,8	31,2	14,7	32,6	32,6	20,2

Таблица 2 – Статистика выбора программ и результативности лечения

Выбор лечебной программы, %	Снижение интенсивности курения и повышение мотивации отказа от курения	6,9
	Короткая программа	40,4
	Длительная программа	52,8
Результаты лечения, %	Отказ от лечебной программы	30,7
	Снижена интенсивность курения	51,4
	Полный отказ от курения	17,9

Корреляционный анализ регистрируемых факторов и результативного признака выборочных данных полнофакторной модели представлен в виде таблицы коэффициентов корреляции (табл. 3).

Таблица 3 – Коэффициенты корреляции полнофакторной модели

r_{ij}	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Y
Z1	0,025	-0,110	-0,324	-0,235	-0,087	-0,081	-0,131	0,025
Z2	-0,108	1,000	0,158	0,211	0,032	0,119	-0,049	-0,108
Z3	-0,324	0,158	1,000	0,728	0,007	0,476	0,037	-0,139
Z4	-0,235	0,211	0,728	1,000	0,136	0,501	0,189	-0,117
Z5	-0,087	0,032	0,007	0,136	1,000	0,102	0,724	0,155
Z6	-0,081	0,119	0,476	0,501	0,102	1,000	0,154	-0,023
Z7	-0,131	-0,049	0,037	0,189	0,724	0,154	1,000	0,149
Y	0,025	-0,108	-0,139	-0,117	0,155	-0,023	0,149	1,000

Анализируя данные корреляционной таблицы, можно отметить повышенную корреляцию таких факторов как Z3-Z4, Z3-Z5, Z4-Z6, Z5-Z7, что говорит о невозможности избежать мультиколлинеарности

полнофакторной модели, приводящей к потере идентифицируемости такой модели, поэтому предлагается в парах факторов с повышенной корреляцией, отказаться от тех факторов, которые в наименьшей степени коррелируют с результативным признаком. Сокращенная статистическая модель, построенная таким образом и свободная от мультиколлинеарности представляется следующим образом

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3, \text{ где } X_1 = Z2, X_2 = Z3, X_3 = Z5. \quad (2)$$

Помимо устранения мультиколлинеарности, необходимого для построения идентифицируемой модели, можно прийти и к чисто практическому выводу о том, что сложившаяся практика и используемая методика назначения программ лечения никотиновой зависимости имеют некоторую избыточность, поскольку регистрация таких факторов как Z2 – возраст, Z3 – индекс курильщика и Z5 – степень мотивации бросить курить является определяющей, а другие факторы не приносят новой статистической информации.

Более того, выявлен весьма интересный результат: низкая корреляция выбранной программы лечения и результатов лечения. Частичное объяснение этому достаточно парадоксальному выводу заключается в том, что врач не выставляет пациенту "нулевой" программы, т.е. "ничего не делать" даже в тех случаях, когда очевидно, что заядлый курильщик никак не будет реагировать на предлагаемую программу. В этих случаях врач всё же назначает некоторую программу, исходя из гуманитарно-этических соображений, что и приводит к описанному парадоксу.

В первом приближении можно считать, что зависимость (2) отражает линейную регрессию и её коэффициенты можно получить методом наименьших квадратов. Для используемых выборочных данных при таком подходе получим следующие оценки

$$b_0^* = 1,854817; b_1^* = -0,004376; b_2^* = -0,000852; b_3^* = 0,064257. \quad (3)$$

Однако, прогностические возможности линейной регрессии в данном случае весьма слабые, что можно понять, имея в виду низкий коэффициент детерминации $R^2 < 0,20$. Тем не менее, решение линейного регрессионного анализа можно использовать в качестве начального приближения, более обоснованного в данной постановке задачи пробит-метода для случая

равноинтервальной дискретной оценки результирующего признака [1].

При таком подходе оценки коэффициентов зависимости (2) находятся с использованием принципа максимума правдоподобия, в соответствии с которым максимизируется функция правдоподобия

$$\ln(b) = \sum_{j=1}^3 \sum_{i: y_i=1} \ln(P(y_i = j / x_i, b)) \rightarrow \max; b = \{b_0; b_1; b_2; b_3\} - \text{var.} \quad (4)$$

При записи функции правдоподобия используются вероятности принятия соответствующих лечебных программ для $i - \tilde{a}_i$ представителя выборочных данных, вычисляемые в соответствии с нормальным законом распределения следующим образом

$$\begin{aligned} P(y_i = 1 / x_i, b) &= F_{\text{норм}}(y_i^{\text{пасч}}; m_{y,1}; s_{y,1}), \text{ если } y_i^{\text{пасч}} < y_{0,1}; \\ P(y_i = 2 / x_i, b) &= F_{\text{норм}}(y_i^{\text{пасч}}; m_{y,2}; s_{y,2}) - P(y_i = 1 / x_i, b), \text{ если } y_{0,1} \leq y_i^{\text{пасч}} < y_{0,2}; \\ P(y_i = 3 / x_i, b) &= 1 - P(y_i = 1 / x_i, b) - P(y_i = 2 / x_i, b), \text{ если } y_i^{\text{пасч}} \geq y_{0,2}. \end{aligned} \quad (5)$$

В соотношениях (3) использованы обозначения

$$\begin{aligned} y_i^{\text{пасч}} &= b_0 + b_1 x_{1,i} + b_2 x_{2,i} + b_3 x_{3,i}, i = 1, \dots, n; \\ y_{0,j} &= F_{\text{норм}}(y_i^{\text{пасч}}, m_{0,j}, s_{0,j}), j = 1, 2, 3; \\ m_{0,j} &= \frac{1}{n_j} \sum_{i:j=j} y_i^{\text{пасч}}, j = 1, 2, 3; \quad s_{0,j} = \sqrt{\frac{1}{n_j} \sum_{i:j=j} (y_i^{\text{пасч}} - m_{0,j})^2}, j = 1, 2, 3. \end{aligned} \quad (6)$$

Численный поиск решения оптимизационной задачи в виде (2) с учётом соотношений (3) и (4) привёл к тем же значениям, которые задавались в качестве начального приближения, что можно считать случайным совпадением т.е.

$$b_0 = 1,854817; b_1 = -0,004376; b_2 = -0,000852; b_3 = 0,064257. \quad (7)$$

Получив оценки вектора $b = \{b_0; b_1; b_2; b_3\}$, можно для любого i -го наблюдения найти вероятности продуктивной реализации каждой из предлагаемых программ $P(y_i = 1 / x_i, b)$, $P(y_i = 2 / x_i, b)$, $P(y_i = 3 / x_i, b)$, используя соотношения (5), что является дополнительным инструментом выбора лечебной программы и последующего обоснования её материального обеспечения.

Для реализации предлагаемой методики разработано программное

обеспечение в редакторе электронных таблиц Excel.

Параметры, используемые в соотношениях (5), рассчитанные на основе соотношений (6) и имеющихся выборочных данных, оцениваются следующим образом:

$$y_{0,1} = 1,447; \quad y_{0,2} = 5,681; \quad m_{0,1} = 3,537; \quad s_{0,1} = 3,238;$$
$$m_{0,2} = 2,382; \quad s_{0,2} = 2,989; \quad m_{0,3} = 3,542; \quad s_{0,3} = 2,081.$$

Для иллюстрации информативности предлагаемых методик прогнозирования результативности назначаемых программ лечения никотиновой зависимости обратимся к примерам, приведенным в табл. 3. Все эти примеры соответствуют реальным выборочным данным с i -ми порядковыми номерами.

Таблица 3 – Примеры прогнозирования лечения

i	$y_i^{\text{расч}}$	Вероятности результативности программ лечения		
		P(1)	P(2)	P(3)
12	4,821	0,0524	0,6078	0,3398
15	1,865	0,4204	0,5463	0,0333
46	-1,193	0,8977	0,1018	0,0005
48	6,717	0,0057	0,3035	0,6908

Обратим внимание на пациентов с номерами $i=12$, $i=15$. В обоих случаях наибольшую вероятность положительной реализации имеет программа № 2, приводящая к снижению интенсивности курения и повышению мотивации отказа от курения. Однако, пациент с номером $i=12$ имеет существенную возможность вообще отказаться от курения, а пациент с номером $i=15$ склонен в значительной мере оказаться от лечения. Очевидно, что назначая этим больным одну и ту же программу лечения, врач должен учитывать возможное несовпадение с ожидаемыми результатами в худшую и в лучшую стороны. В тоже время у пациентов с номерами $i=46$, $i=48$ можно ожидать очевидный безальтернативный результат лечения.

Возвращаясь к выполненному исследованию мультиколлинеарности исходной полнофакторной модели, можно отметить её избыточность и необходимость перехода к усечённой модели (2), что позволяет значительно

увеличить эффективность диагностики за счёт сокращения времени идентификации и регистрации ненужных в статистическом смысле факторов.

Список литературы:

1. Елисеева, И.И. и др. Эконометрика: Учебник / Под редакцией Елисеевой И.И. – М.: Финансы и статистика, 2005. – 576 с.