

БИНАРНАЯ ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ КАК МОДЕЛЬ УПРАВЛЕНИЯ НА ПРИМЕРЕ ЗАДАЧИ ПОВЫШЕНИЯ КАЧЕСТВА ЖИЗНИ НАСЕЛЕНИЯ

^{1,3}Тырсин А.Н., ²Васильева Е.В.

¹ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», Екатеринбург, e-mail: at2001@yandex.ru;

²ФГБУН «Институт экономики» УрО РАН, Екатеринбург, e-mail: elvitvas@ya.ru;

³ФГБУН «Научно-инженерный центр «Надежность и ресурс больших систем и машин» УрО РАН, Екатеринбург, e-mail: at2001@yandex.ru

Бинарная логистическая регрессия – известный метод многомерного статистического анализа, используемый для классификации данных. В экономических задачах возникают ситуации, когда необходимо расширить ее область применения и использовать в задачах мониторинга и выработки управлений решений. Для количественной характеристики качества жизни населения использован широко распространенный показатель – индекс человеческого развития. Целью статьи является разработка инструментария для использования бинарной логистической регрессии в задачах диагностики и управления многомерными системами и его апробация применительно к вопросам увеличения индекса человеческого развития регионов России. Описана математическая модель мониторинга и управления многомерными стохастическими системами, построенная на основе бинарной логистической регрессии. Бинарная логистическая регрессия как модель мониторинга и управления апробирована на примере задачи повышения качества жизни регионов. В качестве вектора исходных данных использован набор социально-экономических показателей регионов России. Описан подход к выработке управлений решений для региона в виде оптимизационных задач увеличения вероятности отнесения его к классу регионов с высоким уровнем качества жизни. На основе выявленной взаимосвязи между качеством жизни населения и социально-экономическими показателями региона решена задача оптимизации.

Ключевые слова: бинарная логистическая регрессия, оптимизационная задача, качество жизни, индекс человеческого развития, социально-экономические показатели, регион

A BINARY LOGISTIC REGRESSION AS A MANAGEMENT MODEL ON THE EXAMPLE OF IMPROVING THE QUALITY OF LIFE OF THE POPULATION

^{1,3}Tyrsin A.N., ²Vasilyeva E.V.

¹Federal State Autonomous Educational Institution of Higher Education Ural Federal University named after the first President of Russia B.N. Yeltsin, Yekaterinburg, e-mail: at2001@yandex.ru;

²Federal State Budgetary Institution of Science Institute of Economics, Ural Branch, Russian Academy of Science, Yekaterinburg, e-mail: elvitvas@ya.ru;

³Federal State Budgetary Institution of Science Scientific-Engineering Center Reliability and Life of Large Systems and Machines, Ural Branch, Russian Academy of Science, Yekaterinburg, e-mail: at2001@yandex.ru

Binary logistic regression is a well-known method of multivariate statistical analysis used to classify data. In economic problems, there are situations when it is necessary to expand its scope and use it in the tasks of monitoring and developing management decisions. Such a widespread indicator as the Human Development Index is used to quantify the quality of life of the population. The purpose of the article is to develop tools for using binary logistic regression in the problems of diagnostics and management of multidimensional systems and to test it in relation to increasing the human development index of Russian regions. A mathematical model for monitoring and managing multidimensional stochastic systems based on binary logistic regression is described. Binary logistic regression as a monitoring and management model is tested on the example of improving the quality of life of regions. A set of socio-economic indicators of Russian regions is used as a vector of initial data. An approach to the development of management decisions for a region is described in the form of optimization tasks to increase the probability of classifying it as a region with a high level of quality of life. In the paper the problem of optimization is solved on the basis of the revealed correlation between the life quality of the population and socio-economic indicators of the region.

Keywords: binary logical regression, optimization problem, quality of life, human development index, socio-economic indicators, region

Бинарная логистическая регрессия – известный метод многомерного статистического анализа, используемый для классификации данных в экономике [1; 2]. Она осуществляет разделение множества многомерных данных линейной границей

на соответствующие заданным классам две области и прогнозирует вероятность соответствия каждой точки этим классам [3].

Достаточно часто, особенно в малоизученных приложениях, возникают ситуации, когда не удается построить качествен-

ную линейную регрессионную зависимость даже с учетом использования замен переменных. Основная причина этого состоит в неполноте модели из-за отсутствия части существенных переменных. Также существенных переменных может оказаться очень много, но они проявляются лишь эпизодически при наступлении некоторых условий и в регрессионной модели будут статистически не значимыми.

В этой ситуации причинно-следственную математическую модель можно построить на основе логистической регрессии, которая не накладывает жестких условий на переменные. Здесь лишь требуется суметь разделить наблюдения на две группы, что является значительно более простой процедурой. Затруднение, связанное с непрерывностью зависимой переменной Y , преодолимо с помощью формирования значений новой бинарной величины

$$u_i = \begin{cases} -1, & y_i < (>)y^*, \\ 1, & y_i \geq (\leq)y^*, \end{cases} \text{ где } y^* - \text{ некоторое пороговое значение.}$$

Но применение логистической регрессии только как классификатора будет недостаточным, необходим инструментарий для ее использования для диагностики и управления.

Рассмотрим иное возможное использование бинарной логистической регрессии как замены в некотором смысле модели линейной регрессии. Целью статьи является разработка инструментария для использования бинарной логистической регрессии в задачах мониторинга и выработки управленических решений многомерными стохастическими системами и его апробация применительно к вопросам увеличения индекса человеческого развития регионов России. Исходные данные взяты из справочников Росстата.

Материалы и методы исследования

Дана выборка прецедентов (обучающая выборка) значений признаков X_1, X_2, \dots, X_m

$$(\mathbf{x}_i, y_i), i = 1, 2, \dots, n, \quad (1)$$

где $\mathbf{x}_i = \begin{pmatrix} x_{i0} \\ x_{i1} \\ \dots \\ x_{im} \end{pmatrix} = \begin{pmatrix} 1 \\ x_{i1} \\ \dots \\ x_{im} \end{pmatrix}$ – вектор значений i -го

объекта, $\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{pmatrix}; \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$,

$y_i \in \{-1; 1\}$ – указывающая на принадлежность i -го объекта соответствующему классу бинарная переменная (для первого класса $y_i = -1$, для второго – $y_i = 1$); m – количество признаков у объектов; n – число наблюдений.

Классификация выполняется с помощью логистической функции [3]

$$h(\mathbf{x}) = \frac{1}{1 + \exp\{-\mathbf{b}^T \mathbf{x}\}}, \quad (2)$$

причем $\forall \mathbf{x} \quad 0 < h(\mathbf{x}) < 1$. Вектор $\mathbf{b}^T = (b_0 \ b_1 \ \dots \ b_m)$ является набором коэффициентов, задающих разделяющую границу, которая представляет собой гиперплоскость

$$\Pi: \mathbf{b}^T \mathbf{x} = 0. \quad (3)$$

Введем функцию $W(\mathbf{x}) = \mathbf{b}^T \mathbf{x}$. Зададим область D_1 возможных значений \mathbf{x} для первого класса как $D_1 = \{\mathbf{x}: W(\mathbf{x}) < 0\}$, а для второго класса – как $D_2 = \{\mathbf{x}: W(\mathbf{x}) > 0\}$. Тогда $\forall \mathbf{x} \in D_1 \quad h(\mathbf{x}) < 0,5, \quad \forall \mathbf{x} \in D_2 \quad h(\mathbf{x}) > 0,5, \quad \forall \mathbf{x} \in \Pi \quad h(\mathbf{x}) = 0,5$ и $\forall \mathbf{x}^* \quad P(\mathbf{x}^* \in D_1) = 1 - h(\mathbf{x}^*)$ и $P(\mathbf{x}^* \in D_2) = h(\mathbf{x}^*)$. Если $\mathbf{x} \in D_1$, то $y = -1$, в противном случае $y = 1$.

В [4] предложен метод вычисления коэффициентов вектора \mathbf{b} с помощью решения задачи

$$Q(\mathbf{b}) = \sum_{i=1}^n \ln \left(1 + e^{-y_i \mathbf{b}^T \mathbf{x}_i} \right) \rightarrow \min_{\mathbf{b} \in \mathbf{R}^m}. \quad (4)$$

Вектор коэффициентов \mathbf{b} оценивают разными алгоритмами, например используют алгоритм Ньютона–Рафсона [4–6]. Однако при корректной классификации всех наблюдений целевая функция $Q(\mathbf{b})$ имеет нулевую нижнюю грань на бесконечности, и поэтому задача (4) не будет иметь точного решения. Увеличение компонент вектора \mathbf{b} вызывает неограниченный рост некоторых значений $-y_i \mathbf{b}^T \mathbf{x}_i$. В результате этого происходит рост вычислительных погрешностей, приводящий к переполнению памяти и к остановке алгоритма [7].

В [8] описан алгоритм нулевого порядка, использующий на каждой итерации случайный поиск с фиксацией длины вектора \mathbf{b} . Это позволяет устранить неконтролируемый рост вычислительных погрешностей и обеспечивает устойчивость работы алгоритма.

Рассмотрим задачу классификации регионов по качеству жизни населения на два класса. Используем данные за 2013 г., так как это последний год перед усложнени-

ем внешнеполитической (введение санкций против России из-за присоединения Крыма) и внешнеэкономической (падение цен на нефть) обстановки. Для численной оценки качества жизни населения регионов используем индекс человеческого развития (ИЧР) [9]. В качестве компонент вектора \mathbf{x} будем использовать набор социально-экономических показателей регионов России. Регионы разделим на два класса, к первому классу D_1 отнесем регионы с низким ИЧР ($y = -1$), а ко второму классу D_2 с высоким ИЧР ($y = 1$). Это разделение описано в [10].

Будем использовать вероятность $P(\mathbf{x} \in D_2) = h(\mathbf{x})$ отнесения региона к классу D_2 как целевую функцию. Фактически это означает, что мы стремимся повысить качество жизни населения региона посредством выработки и реализации управлений рекомендаций, повышающих вероятность $h(\mathbf{x})$ его отнесения к классу регионов с высоким ИЧР.

Как показано в [10], вероятность $h(\mathbf{x})$ зависит от расстояния $d = |\mathbf{b}^T \mathbf{x}| / \sqrt{\sum_{j=1}^m b_j^2}$ от гиперплоскости (3) до точки \mathbf{x} до и градиента $\text{grad } h(\mathbf{x})$ и вектора-градиента

$$\begin{aligned} \text{grad } h(\mathbf{x}) &= \left(\frac{\partial h(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial h(\mathbf{x})}{\partial x_m} \right) = \\ &= \frac{\exp\{-\mathbf{b}^T \mathbf{x}\}}{(1 + \exp\{-\mathbf{b}^T \mathbf{x}\})^2} \cdot (b_1, \dots, b_m). \end{aligned}$$

Поскольку в (3) вектор коэффициентов \mathbf{b} задан с точностью до произвольного множителя, то введем нормировку: $\mathbf{b}^0 = \mathbf{b} / |\mathbf{b}|$, $|\mathbf{b}^0| = \sum_{j=1}^m (b_j^0)^2 = 1$.

$$\text{Считаем } Z(\mathbf{x}) = \frac{h(\mathbf{x})}{1 - h(\mathbf{x})} = \frac{P(\mathbf{x} \in D_2)}{P(\mathbf{x} \in D_1)},$$

$$\text{т.е. } \forall k \ln Z(\mathbf{x}) = b_0 + \sum_{j=1}^m b_j x_j, b_k = \frac{d_{x_k} Z(\mathbf{x})}{Z(\mathbf{x}) dx_k}.$$

Если приравнять $dx_k = \Delta x_k = 1$, то последняя формула в процентах примет вид:

$$b_k \cdot 100\% \approx \frac{\Delta_{x_k} Z(\mathbf{x})}{Z(\mathbf{x})} \cdot 100\%. \text{ То есть значение } b_k$$

показывает, на сколько процентов изменится отношение $P(\mathbf{x} \in D_2)$ к $P(\mathbf{x} \in D_1)$ при увеличении X_k на единицу при фиксации остальных показателей.

Производная функции $Z(\mathbf{x})$ по X_k прямо пропорциональна коэффициенту b_k . Рост (уменьшение) функции $Z(\mathbf{x})$ равносителен увеличению (уменьшению) $h(\mathbf{x})$. Следовательно, для увеличения значения

$h(\mathbf{x})$ приращения Δx_k должны быть положительными для $b_k > 0$ и отрицательными, если $b_k < 0$.

По статистическим данным за 2013 г. были сформированы два класса регионов [10]. В класс D_1 были включены регионы с уровнем ИЧР не выше 0,84. В класс D_2 попали регионы с уровнем ИЧР не менее 0,85. С целью повышения достоверности результатов были удалены из рассмотрения регионы с очень низким и очень высоким уровнем ИЧР и с пропусками в данных. Используемые социально-экономические показатели и рассчитанные коэффициенты уравнения (3) для классификации регионов по ИЧР указаны в табл. 1.

Из табл. 1 видим, что для увеличения вероятности $h(\mathbf{x})$ (повышения качества жизни в регионе) необходимо:

– уменьшать значения показателей $X_1, X_6, X_7, X_9, X_{10}, X_{12}, X_{13}$ (коэффициенты b_i отрицательные),

– увеличивать значения показателей $X_2, X_3, X_4, X_5, X_8, X_{11}$ (коэффициенты b_i положительные).

Отметим, что максимальное значение коэффициента $b_4 = 10,004$ объясняется кейнсианской теорией [11].

Коэффициенты уравнения (2) были определены по данным за 2013 г. (табл. 1). Исследуем адекватность модели на статистических данных за следующий 2014 г. на примере областей Уральского федерального округа. В табл. 2 приведены фактические значения ИЧР и оценки вероятностей $h(\mathbf{x})$ за 2013–2014 гг. Отметим, что в 2013 г. вероятности $h(\mathbf{x})$ находились по обучающим выборкам, а в 2014 г. – по имеющимся тестовым статистическим данным по формуле (2). При расчете количественных значений признаков X_4, X_5, X_{12}, X_{13} за 2014 г. денежные показатели пересчитывались с учетом инфляции, оцениваемой через индекс потребительских цен.

Тюменская область имеет в обоих случаях наибольшее значение ИЧР, что соответствует максимальной вероятности $h(\mathbf{x})$, практически равной 1. Росту значений ИЧР в остальных областях соответствует увеличение для всех трех регионов вероятностей $h(\mathbf{x})$.

Результаты исследования и их обсуждение

В [10] был сформулирован подход к управлению регионом в виде оптимизационных задач повышения качества жизни посредством увеличения вероятности $h(\mathbf{x})$ отнесения его к классу регионов с высоким ИЧР. Рассмотрим эти задачи. Первая задача – это максимизация вероятности $h(\mathbf{x})$ при

ограничениях на изменения социально-экономических показателей

$$\begin{cases} h(\mathbf{x}) \rightarrow \max, \\ x_j = x_j^0 + \Delta_j, j = 1, \dots, m, \\ \Delta_j \in G_j, j = 1, \dots, m, \end{cases} \quad (5)$$

где \mathbf{x}^0 – вектор значений показателей; G_j и Δ_j – область допустимых значений изменений и изменение компоненты X_j соответственно.

В задаче (5) $\text{grad } h(\mathbf{x})$ всегда ортогонален гиперплоскости (2) и не учитываются экономические ограничения и затраты, необходимые для изменения социально-экономических показателей. Поэтому ее реше-

нием будет точка пересечения вектора $\overrightarrow{\mathbf{x}^0 \mathbf{x}^1}$ ($\mathbf{x}^1 = \mathbf{x}^0 + a\mathbf{b}$, $a > 0$) с границей допустимой области.

Ниже учтем экономические ограничения и затраты на изменения X_j

$$\begin{cases} h(\mathbf{x}) \rightarrow \max, \\ x_j = x_j^0 + \Delta_j, j = 1, \dots, m, \\ \Delta_j \in G_j, j = 1, \dots, m, \\ v_j(\Delta_j) \leq V_j, j = 1, \dots, m, \end{cases} \quad (6)$$

где V_j и $v_j(\Delta_j)$ – предельная величина затрат и функция затрат на изменение X_j соответственно.

Показатели и коэффициенты уравнения (3)

| Показатель | Обозначение | b_j |
|--|-------------|--------|
| Свободный член, 1 | X_0 | 0,072 |
| Суммарный коэффициент рождаемости, ед. | X_1 | -2,381 |
| Ожидаемая продолжительность жизни при рождении, лет/100 | X_2 | 0,807 |
| Количество ипотечных жилищных кредитов, предоставленных кредитными организациями физическим лицам-резидентам, ед. на 1000 чел. | X_3 | 0,879 |
| Инвестиции в основной капитал на душу населения в фактически действовавших ценах, руб./ 10^5 | X_4 | 10,004 |
| Объем валового регионального продукта на душу населения, млн руб. на чел. | X_5 | 3,974 |
| Отношение объема инвестиций в основной капитал к валовому региональному продукту, ед.*10 | X_6 | -1,911 |
| Смертность населения от новообразований, число умерших на 1000 чел. | X_7 | -1,149 |
| Отношение средней заработной платы младшего медицинского персонала к средней заработной плате по субъекту РФ, ед.*10 | X_8 | 0,165 |
| Отношение средней заработной платы среднего медицинского персонала к средней заработной плате по субъекту РФ, ед. | X_9 | -1,770 |
| Уровень общей безработицы, %/10 | X_{10} | -2,739 |
| Обеспеченность населения врачами, чел. на 1000 чел. | X_{11} | 0,688 |
| Консолидированный бюджет субъекта РФ и территориального государственного внебюджетного фонда на ЖКХ, тыс. руб. на 10 чел. | X_{12} | -4,223 |
| Консолидированный бюджет субъекта РФ и территориального внебюджетного фонда на здравоохранение, тыс. руб. на 10 чел. | X_{13} | -0,870 |

Таблица 1

Фактические значения ИЧР и оценки вероятностей $h(\mathbf{x})$ отнесения субъектов УрФО к группе регионов с высоким качеством жизни за 2013–2014 гг.

| Показатель | Год | Курганская область | Свердловская область | Тюменская область | Челябинская область |
|-----------------|------|--------------------|----------------------|-------------------|---------------------|
| ИЧР | 2013 | 0,829 | 0,868 | 0,901 | 0,848 |
| | 2014 | 0,831 | 0,873 | 0,903 | 0,857 |
| $h(\mathbf{x})$ | 2013 | 0,005 | 0,851 | 1,000 | 0,805 |
| | 2014 | 0,015 | 0,936 | 1,000 | 0,884 |

Достижение функцией $h(\mathbf{x})$ заданной вероятности p_0 при минимальных затратах на изменение вектора \mathbf{x} можно представить в виде задачи

$$\begin{cases} \sum_{j=1}^m v_j(\Delta_j) \rightarrow \min, \\ x_j = x_j^0 + \Delta_j, j = 1, \dots, m, \\ \Delta_j \in G_j, j = 1, \dots, m, \\ h(\mathbf{x}) = p_0. \end{cases} \quad (7)$$

Значение целевой функции в задачах (5)–(7) позволяет оценить прямые последствия реализации заданного воздействия на объект, а косвенные – в каждом конкретном случае можно определить на основе полученного в результате решения оптимизационной задачи значения вектора \mathbf{x} .

Отметим, что в задачах (6) и (7) достаточно сложно задавать функции затрат $v_j(\Delta_j)$. Обойдем это ограничение, сформулировав задачу следующим образом:

$$\begin{cases} \sum_{j=1}^m \rho_j \cdot (x_j / x_j^0 - 1)^2 \rightarrow \min, \\ h(\mathbf{x}) = p_0, \\ x_j, \rho_j \geq 0, j = 1, \dots, m. \end{cases} \quad (8)$$

Здесь решается вопрос перехода региона в состояние $h(\mathbf{x}) = p_0$ при минимальном среднеквадратическом относительном изменении значений компонент вектора социально-экономических показателей. Такая постановка не требует задавать в явном виде трудоемкость изменения показателей X_j . Весовые коэффициенты ρ_j позволяют учесть особенности, связанные с изменением переменных. Если априорная информация о показателях X_j , включенных в (2), отсутствует, то считаем все ρ_j равными 1.

Нулевое значение весового коэффициента можно использовать в ситуациях, когда знак коэффициента b_j имеет противоположное направление по сравнению с желаемым, позволяет зафиксировать фактическое значение показателя X_j . Например, здесь $b_1 = -2,381 < 0$, данная ситуация вызвана имеющейся закономерностью, связанной с тем, что в регионах с более высоким качеством жизни рождаемость ниже по сравнению с регионами с низким качеством жизни. Это в [12] названо обратной связью рождаемости с уровнем жизни. Поскольку вектор \mathbf{b} определяется по обучающей выборке и отражает сложившуюся социально-экономическую ситуацию

в стране, данный вопрос не может решаться в рамках рассматриваемых оптимизационных задач.

Введя в задачу (8) дополнительные ограничения на предельные изменения компонент, учитывающие особенности изменения социально-экономических показателей X_j , получим

$$\begin{cases} \sum_{j=1}^m \rho_j \cdot (x_j / x_j^0 - 1)^2 \rightarrow \min, \\ h(\mathbf{x}) = p_0, \\ x_j = x_j^0 + \Delta_j, \Delta_j \in G_j, j = 1, \dots, m, \\ x_j \geq 0, \rho_j \geq 0, j = 1, \dots, m. \end{cases} \quad (9)$$

Апробируем логистическую регрессию как модель управления (8) на примере Курганской, Свердловской и Челябинской областей, считая $\rho_j = 1, j = 1, 2, \dots, m$.

Курганская область относится к регионам с низким уровнем ИЧР. Вероятность отнесения ее к регионам с высоким уровнем ИЧР: $h(\mathbf{x}^0) = 0,005$. В табл. 3 представлены результаты управления с целью выхода Курганской области на пороговый уровень ($h(\mathbf{x}^*) = 0,5$), полученные с помощью решения задачи (8). Для сопоставления расчетных значений показателей в табл. 3 приведены фактические значения за 2013 г. (год исходных значений для проведения расчетов) и 2016 г., а также целевые значения показателей, взятые из указов Президента РФ № 596-600 и № 606 от 07.05.2012.

Расчетные (оптимальные) значения социально-экономических показателей показывают тот минимальный уровень, достижение которого позволяет субъекту РФ повысить значение ИЧР. Поэтому расчетные значения показателей не могут выступать в качестве новых предлагаемых целевых ориентиров, но их следует принимать во внимание при оптимизации затрат и уточнении приоритетов дальнейшего развития.

Для Курганской области приоритетами в повышении качества жизни должны стать: а) увеличение продолжительности жизни населения (X_2) до 70 лет (этот уровень в целом по России был достигнут еще в 2012 г., но для данной области при изначально его невысоком значении является оптимальным и реалистичным ориентиром); б) развитие ипотечного жилищного кредитования (X_3), объемы которого в последние годы были серьезно сокращены в результате макроэкономической нестабильности; в) рост инвестиционной активности (X_4) позволит увеличить производство, а также привлечь свободную рабочую силу и повысить пла-

тежеспособность населения; г) снижение уровня безработицы (X_{10}) до 6,7%.

Свердловская область обладает относительно высоким уровнем ИЧР. Результаты решения оптимизационной задачи, в которую было заложено достижение более высокого уровня ИЧР, соответствующее увеличению вероятности $h(\mathbf{x})$ до 0,95, представлены в табл. 4.

Важным условием увеличения вероятности роста ИЧР в Свердловской области является экономический рост, невозможный без дополнительных инвестиций (X_4). Развитие же ипотечного жилищного кредитования (X_3) способствует привлечению инвестиций в строительную отрасль – драйвера экономического роста территории.

Таблица 3
Фактические, целевые и расчетные показатели Курганской области

| Показатель | Фактические значения за 2013 г. (\mathbf{X}^0) | Фактические значения за 2016 г. | Целевые значения | Расчетные значения (\mathbf{X}^*) |
|------------|--|---------------------------------|------------------|---------------------------------------|
| X_1 | 2,12 | 2,03 | 1,753 | 1,55 |
| X_2 | 0,68 | 0,69 | 0,74 | 0,70 |
| X_3 | 5,33 | 5,44 | 5,69** | 6,66 |
| X_4 | 0,38 | 0,33 | – | 0,45 |
| X_5 | 0,19 | 0,21* | – | 0,19 |
| X_6 | 1,99 | 1,54* | 2,7 | 1,60 |
| X_7 | 2,71 | 2,58 | 1,93 | 2,26 |
| X_8 | 4,64 | 5,5 | 10,0 | 4,83 |
| X_9 | 0,92 | 1,03 | 1,00 | 0,84 |
| X_{10} | 0,75 | 0,84 | – | 0,67 |
| X_{11} | 2,60 | 2,88* | – | 2,85 |
| X_{12} | 0,21 | 0,24 | – | 0,20 |
| X_{13} | 1,28 | 1,75 | – | 1,21 |

П р и м е ч а н и е : * – за 2015 г.; ** – значение рассчитано, исходя из установленного целевого значения количества выдаваемых ипотечных жилищных кредитов в размере 815 тысяч в год в целом по России при сохранении численности населения на уровне 2012 г.

Таблица 4
Фактические, целевые и расчетные показатели Свердловской области

| Показатель | Фактические значения за 2013 г. (\mathbf{X}^0) | Фактические значения за 2016 г. | Целевые значения | Расчетные значения (\mathbf{X}^*) |
|------------|--|---------------------------------|------------------|---------------------------------------|
| X_1 | 1,87 | 1,91 | 1,753 | 1,807 |
| X_2 | 0,698 | 0,70 | 0,74 | 0,701 |
| X_3 | 6,153 | 6,22 | 5,69** | 6,403 |
| X_4 | 0,817 | 0,799 | – | 0,868 |
| X_5 | 0,363 | 0,411* | – | 0,371 |
| X_6 | 2,25 | 1,97* | 2,7 | 2,154 |
| X_7 | 2,239 | 2,219 | 1,93 | 2,196 |
| X_8 | 5,28 | 6,1 | 10,0 | 5,315 |
| X_9 | 0,927 | 0,97 | 1,00 | 0,916 |
| X_{10} | 5,9 | 6,2 | – | 0,583 |
| X_{11} | 3,66 | 4,24* | – | 3,729 |
| X_{12} | 0,312 | 0,278 | – | 0,308 |
| X_{13} | 1,696 | 1,916 | – | 1,677 |

П р и м е ч а н и е : * – за 2015 г.; ** – значение рассчитано, исходя из установленного целевого значения количества выдаваемых ипотечных жилищных кредитов в размере 815 тысяч в год в целом по России при сохранении численности населения на уровне 2012 г.

Таблица 5

Фактические, целевые и расчетные показатели Челябинской области

| Показатель | Фактические значения за 2013 г. (\mathbf{X}^0) | Фактические значения за 2016 г. | Целевые значения | Расчетные значения (\mathbf{X}^1) | Расчетные значения (\mathbf{X}^2) | Расчетные значения (\mathbf{X}^3) |
|------------|--|---------------------------------|------------------|---------------------------------------|---------------------------------------|---------------------------------------|
| X_1 | 1,8 | 1,84 | 1,753 | 1,78 | 1,76 | 1,72 |
| X_2 | 0,695 | 0,705 | 0,74 | 0,70 | 0,70 | 0,70 |
| X_3 | 8,16 | 7,73 | 5,69** | 8,28 | 8,46 | 8,75 |
| X_4 | 0,62 | 0,55 | — | 0,62 | 0,64 | 0,65 |
| X_5 | 0,25 | 0,33* | — | 0,25 | 0,25 | 0,25 |
| X_6 | 2,44 | 1,86* | 2,7 | 2,42 | 2,39 | 2,33 |
| X_7 | 2,32 | 2,39 | 1,93 | 2,31 | 2,29 | 2,26 |
| X_8 | 4,68 | 5,5 | 10,0 | 4,69 | 4,7 | 4,72 |
| X_9 | 0,78 | 0,90 | 1,00 | 0,78 | 0,78 | 0,77 |
| X_{10} | 0,6 | 0,71 | — | 0,6 | 0,59 | 0,59 |
| X_{11} | 3,67 | 3,94* | — | 3,69 | 3,72 | 3,76 |
| X_{12} | 0,25 | 0,19 | — | 0,25 | 0,25 | 0,25 |
| X_{13} | 1,30 | 1,76 | — | 1,29 | 1,29 | 1,28 |

Причина: * – за 2015 г.; ** – значение рассчитано, исходя из установленного целевого значения количества выданных ипотечных жилищных кредитов в размере 815 тысяч в год в целом по России при сохранении численности населения на уровне 2012 г.

Челябинская область отнесена к регионам со средним уровнем ИЧР с вероятностью $h(\mathbf{x}^0) = 0,805$. При решении оптимизационной задачи повышения ИЧР были заложены три варианта увеличения $h(\mathbf{x})$. Результаты решения для трех вероятностей $h(\mathbf{x}^1) = 0,85$, $h(\mathbf{x}^2) = 0,9$, $h(\mathbf{x}^3) = 0,95$ приведены в табл. 5.

Можно рассматривать разные варианты развития Челябинской области в зависимости от возможностей региона. Чем благоприятнее наблюдаемая динамика по показателям, тем выше вероятность повышения качества жизни в области. Однако по отдельным показателям (X_2 , X_5 , X_{10} , X_{12} и X_{13}) разница в их значениях по трем сценариям незначительна или отсутствует, что говорит об оптимальности данных значений для анализируемой региональной системы. В Челябинской области приоритетом в повышении качества жизни населения должно стать ипотечное кредитование (X_3).

Выводы

1. Показано, что бинарную логистическую регрессию можно использовать как математическую модель в задачах мониторинга и управления многомерными системами.

2. Бинарная логистическая регрессия как модель управления апробирована на примере задачи повышения качества жизни регионов.

Работа выполнена при поддержке РФФИ, грант № 20-41-660008 р_а.

Список литературы

- Попова П.А., Ротмистров А.Н. Логистическая регрессия с категориальными предикторами и эффектами взаимодействия и CHAID: сравнительный анализ на эмпирическом примере // Социология: 4М. 2016. № 43. С. 63–99.
- Кацов И. Машинное обучение для бизнеса и маркетинга. СПб.: Питер. 2019. 512 с.
- Hosmer D.W., Lemeshow S., Sturdivant R.X. Applied Logistic Regression. 3d ed. Wiley. 2013. 397 p.
- Воронцов К.В. Лекции по алгоритмам восстановления регрессии. 2007. [Электронный ресурс]. URL: <http://www.ccas.ru/vorون/download/Regression.pdf> (дата обращения: 15.09.2020).
- Hilbe J.M. Logistic Regression Models. CRC Press. 2009. 656 р.
- Пантелеев А.В., Летова Т.А. Теория оптимизации для инженеров и экономистов. М.: Вузовская книга, 2016. 568 с.
- Васильев Н.П., Егоров А.А. Опыт расчета параметров логистической регрессии методом Ньютона–Рафсона для оценки зимостойкости растений // Математическая биология и биоинформатика. 2011. Т. 6. № 2. С. 190–199. [Электронный ресурс]. URL: http://www.matbio.org/article_pdf.php?id=82 (дата обращения: 15.09.2020).
- Тырсин А.Н., Костин К.К. Оценивание логистической регрессии как экстремальная задача // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2017. № 3(40). С. 52–60. DOI: 10.17223/19988605/40/6.
- Баева О.Н. Индекс человеческого развития: методики определения и оценки на уровне региона // Известия Иркутской государственной экономической академии (Байкальский государственный университет экономики и права). 2012. № 5. С. 143–147.
- Тырсин А.Н., Чистова Е.В., Костин К.К. Моделирование взаимосвязи между качеством жизни и социально-экономическими показателями регионов России // Государственное управление. Электронный вестник. 2016. № 59. С. 212–237.
- Кейнс Дж.М. Общая теория занятости, процента и денег: Пер. с англ. М.: ЗАО «Бизнеском», 2013. 402 с.
- Чурилова Е. Анализ взаимозависимости уровня рождаемости и индекс развития человеческого потенциала. Презентация доклада. Демоскоп Weekly, № 399–400. 23 ноября – 6 декабря 2009. [Электронный ресурс]. URL: <http://www.demoscope.ru/weekly/2009/0399/student02.php> (дата обращения: 15.09.2020).