

МОДЕЛИ МНОЖЕСТВЕННОГО ВЫБОРА В ЗАДАЧАХ ОЦЕНКИ И УПРАВЛЕНИЯ ЭКОЛОГО-ЭКОНОМИЧЕСКИМИ РИСКАМИ

В статье для оценки и управления эколого-экономическими рисками в разрезе административно-территориальных образований Оренбургской области впервые использованы одномерные и двумерные модели множественного выбора. Построен интегральный показатель риска для населения от воздействия окружающей среды и выделены приоритетные объекты управления риском.

Важным этапом управления любыми рисками является их количественная оценка.

На основе данных о среднегодовых концентрациях химических веществ, содержащихся в питьевой воде и/или атмосферном воздухе 35 районов и 4 городов Оренбургской области в 1998-2005 гг.: азот аммиака, алюминий, бор, железо, марганец, медь, молибден, мышьяк, нефть и нефтепродукты, нитраты, свинец, хром (в питьевой воде, пероральное поступление) и диоксид азота, аммиак, взвешенные вещества, оксид углерода, диоксид серы, сероводород, формальдегид (в атмосферном воздухе, ингаляционное поступление), – были рассчитаны оценки рисков возникновения канцерогенных (для веществ, относящихся к канцерогенам по классификации МАИР) и неканцерогенных эффектов для человека от хронического воздействия указанных веществ (данные представлены ФГУЗ «Центр гигиены и эпидемиологии по Оренбургской области»). Так как многие из перечисленных веществ оказывают воздействие на одни и те же системы организма человека, то для характеристики риска развития неканцерогенных эффектов при комбинированном и комплексном воздействии химических соединений был осуществлен переход от коэффициентов опасности HQ к индексам опасности HI для различных систем организма человека [1], [2].

Поскольку города и районы области характеризуются многочисленными рисками, то необходимо решить задачу классификации для выделения наиболее приоритетных групп городов и районов, в отношении которых можно было бы предложить единую стратегию управления. Для классификации использовался метод многомерного неметрического шкалирования, дающий в отличие от кластерного

анализа количественное координатное представление для каждого из объектов, что позволяет анализировать их взаимное расположение. Кроме того, неметрическое шкалирование не требует многомерного нормального распределения признаков, линейности зависимостей между ними, а по сравнению с факторным анализом обычно приводит к извлечению меньшего количества осей (шкал). Было выделено 2 шкалы, величина нормализованного стресса составила от 0,001 до 0,007 для разных лет (реализовано в пакете SPSS). В таблице 1 приведены результаты разбиения городов и районов Оренбургской области за 2005 г., на рисунке 1 – динамика изменения количества объектов в каждом классе за 1998-2005 гг. (в долях от общего количества объектов).

Как видно из рисунка 1, за исследуемый период количество городов и районов Оренбургской области, население которых подвергается высокому и среднему риску от воздействия окружающей среды, существенно увеличилось. При этом наибольшие изменения наблюдаются в динамике класса именно высокого риска: в 1998 году таких городов и районов было 15% от общего числа городов и районов области, а в 2005 стало уже 30%.

Анализируя результаты классификации, можно сделать вывод, что особое внимание следует уделить следующим административно-территориальным образованиям области: городам Медногорск и Орск, Пономаревскому и Соль-Илецкому районам (ежегодно относились к классу высокого риска); городам Оренбург и Новотроицк, Абдулинскому, Беляевскому, Бузулукскому, Грачевскому, Курманаевскому, Матвеевскому, Первомайскому, Светлинскому, Сорочинскому, Тоцкому и Шарлыкскому районам (переход в класс с большей степенью риска).

Таблица 1. Состав классов, выделенных по степени риска для населения от воздействия окружающей среды (2005 г.)

Класс	Количество объектов в классе	Состав класса
Высокий риск	12	Города: Медногорск, Новотроицк, Оренбург, Орск Районы: Абдулинский, Беляевский, Бугурусланский, Бузулукский, Грачевский, Пономаревский, Соль-Илецкий, Тоцкий
Средний риск	13	Районы: Гайский, Кувандыкский, Курманаевский, Матвеевский, Новоорский, Новосергиевский, Оренбургский, Первомайский, Сакмарский, Светлинский, Сорочинский, Шарлыкский, Ясенский
Низкий риск	14	Районы: Адамовский, Акбулакский, Александровский, Асекеевский, Домбаровский, Илекский, Кваркенский, Красногвардейский, Октябрьский, Переволоцкий, Саракташский, Северный, Ташлинский, Тюльганский

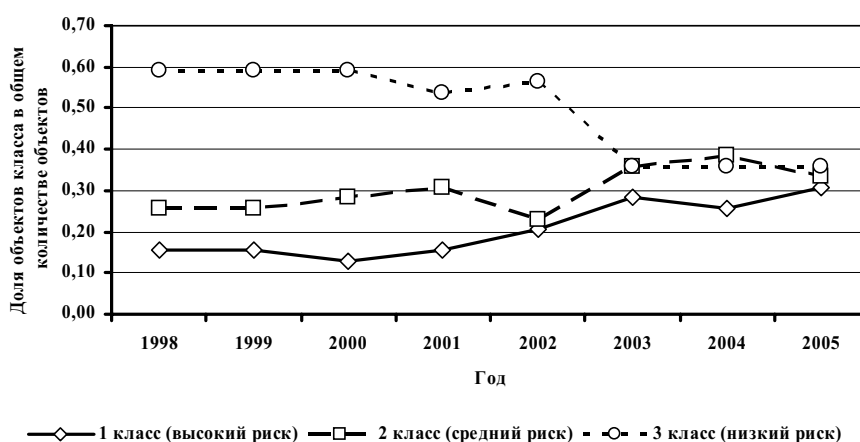


Рисунок 1. Динамика количества объектов в классах за 1998-2005 гг.

Для ранжирования городов и районов области по степени риска и отслеживания изменения их состояния необходимо решить задачу построения интегрального показателя (ИП) риска. Для этого на основе выполненной классификации для характеристики степени риска для населения в i -м городе/районе введем новую переменную, принимающую три значения: 1, если степень риска в городе/районе высокая, 2 – если степень риска средняя и 3 – если степень риска низкая. Построение интегрального показателя y_i^* в ситуации, когда результирующий показатель измерен в порядковой шкале, а объясняющие – в интервальной, может осуществляться на основе порядковых моделей множественного выбора [3], [4].

В моделях такого типа значения искомого интегрального показателя y_i^* связаны со значениями x_{i1}, \dots, x_{ip} объясняющих переменных для i -го объекта как:

$$y_i^* = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i=1, \dots, n, \quad (1)$$

где ε_i – ошибка, отражающая влияние на y_i^* неучтенных факторов. ε_i представляют собой независимо и одинаково распределенные случайные величины, о функции распределения F которых будет сказано ниже.

Хотя переменная y^* латентна, известно, что ее значения связаны со значениями *наблюдаемой* переменной y_i следующим образом:

$$y_i = \begin{cases} 1, & y_i^* \leq \gamma_1, \\ \dots \\ j, & \gamma_{j-1} < y_i^* \leq \gamma_j, \\ \dots \\ k, & y_i^* > \gamma_{k-1} \end{cases}, \quad (2)$$

где $\gamma_1 < \dots < \gamma_j < \dots < \gamma_{k-1}$ – пороговые значения (или границы разделения), вообще говоря, также ненаблюдаемые и подлежащие оцениванию.

Тогда вероятность отнесения объекта i к классу $j \in \{1, \dots, k\}$:

$$P\{y_i = j | x_i\} = P\{\gamma_{j-1} < y_i^* \leq \gamma_j | x_i\} = P\{\gamma_{j-1} < x_i^T \beta + \varepsilon_i \leq \gamma_j\} = P\{\gamma_{j-1} - x_i^T \beta < \varepsilon_i \leq \gamma_j - x_i^T \beta\} = F(\gamma_{j-1} - x_i^T \beta) - F(\gamma_j - x_i^T \beta)$$

Если в качестве функции F рассматривать функцию стандартного нормального распределения $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$, то получаем про-

бит-модель. Неизвестные $p+k-1$ параметров модели (1)-(2) оцениваются методом максимального правдоподобия (ММП). Логарифмическая функция правдоподобия имеет вид:

$$\log L(\beta, \gamma) = \sum_{i=1}^n \sum_{j=1}^k s_{i,j} \log [F(\gamma_{j-1} - x_i^T \beta) - F(\gamma_j - x_i^T \beta)], \quad (3)$$

где

$$s_{i,j} = \begin{cases} 1, & \gamma_{j-1} < y_i^* < \gamma_j \\ 0, & y_i^* \notin (\gamma_{j-1}; \gamma_j) \end{cases}$$

Для моделей типа (1)-(2) особенно важно выполнение предположения о нормальности ошибок ε_i – при его нарушении метод максимального правдоподобия в общем случае не дает состоятельных оценок [5]. В такой ситуации можно использовать, например, «полунепараметрический» метод оценивания («semi-nonparametric», или SNP), предложенный Стюартом [6]. Суть метода заключается в поиске оценок пара-

метров модели путем максимизации функции псевдоправдоподобия, получающейся из (3) заменой функции $F(u)$ на функцию $F_L(u)$:

$$F_L(u) = \frac{\int_{-\infty}^u \left(\sum_{m=0}^M \lambda_m \varepsilon^m \right)^2 \phi(\varepsilon) d\varepsilon}{\int_{-\infty}^{\infty} \left(\sum_{m=0}^M \lambda_m \varepsilon^m \right)^2 \phi(\varepsilon) d\varepsilon}, \quad (4)$$

где $\phi(\varepsilon)$ – функция плотности нормального распределения,

$\lambda = (\lambda_0, \lambda_1, \dots, \lambda_M)$ – вектор дополнительных подлежащих оцениванию так называемых SNP-параметров (весовых коэффициентов),

M – количество SNP-параметров, определяемое путем перебора моделей с различным M на основе использования статистики отношения правдоподобия.

Такое семейство «полунепараметрических» распределений может использоваться для аппроксимации широкого класса распределений, в том числе островершинных с тяжелыми хвостами.

Оценки модели интегрального показателя, полученные в пакете Stata, приведены в таблице 2. Величина псевдо \hat{R}^2 – составила 0,75 и 0,80 для модели, оцененной методом максимального правдоподобия, и для модели, оцененной полунепараметрическим методом, соответственно.

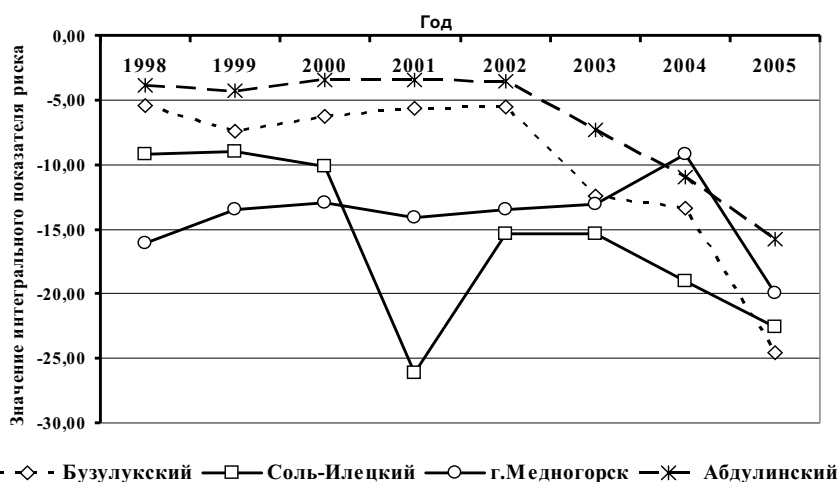


Рисунок 2. Динамика значений интегрального показателя экологического риска для некоторых городов и районов области за 1998-2005 гг.

На основе оцененной модели были получены значения интегрального показателя для каждого из исследуемых городов и районов области за 1998-2005 гг. (частично приведены на рисунке 2 и в таблице 3). Как следует из рисунка 2 и таблицы 3, за период 1998-2005 гг. имел место существенный рост риска для населения (уменьшение значений интегрального показателя).

Одной из наиболее актуальных задач при анализе эколого-экономических рисков является определение вероятности повышенной заболеваемости населения вследствие действия неблагоприятной окружающей среды.

Необходимость моделирования заболеваемости населения обуславливается тем, что рост заболеваемости ведет к возрастанию экономических потерь как на региональном уровне, так и на уровне отдельных предприятий и домохозяйств.

Построим модель, связывающую между собой нагрузку на окружающую среду, риски для населения от воздействия окружающей среды, показатель превышения среднеобластной заболеваемости населения и показатели, характеризующие социально-экономическое положение города или района (данные предоставлены территориальным

Таблица 2. Результаты оценивания порядковой модели, характеризующей риск для населения от воздействия окружающей среды

	Метод оценивания							
	макс. правдоподобия, ММП				«полунепараметрический», SNP			
	оценка	станд. ошибка	z-стат.	p	оценка	станд. ошибка	z-стат.	p
Коэффициенты								
железо	-0,988	0,429	-2,30	0,02	-0,681	0,479	-1,42	0,16
мышьяк	-200,080	48,156	-4,15	0,00	-179,722	27,057	-6,64	0,00
нитраты	-0,029	0,011	-2,57	0,01	-0,017	0,009	-1,92	0,05
свинец	-2,377	2,001	-1,19	0,24	-1,903	3,049	-0,62	0,53
хром	-131,081	19,194	-6,83	0,00	-114,012	14,257	-8,00	0,00
медь	-2,420	0,937	-2,58	0,01	-1,699	0,931	-1,82	0,07
марганец	-5,644	2,559	-2,21	0,03	-3,750	3,914	-0,96	0,33
взвеш.в-ва	-16,152	2,474	-6,52	0,00	-15,251	0,842	-18,11	0,00
диоксид азота	-16,230	4,361	-3,72	0,00	-14,055	2,408	-5,84	0,00
сероводород	-447,308	102,105	-4,38	0,00	-389,725	83,003	-4,70	0,00
оксид углерода	-0,251	0,068	-3,69	0,00	-0,178	0,064	-2,76	0,01
формальдегид	-75,049	30,517	-2,46	0,01	-84,642	7,898	-10,72	0,00
аммиак	-6,851	1,093	-6,27	0,00	-5,919	0,448	-12,99	0,00
Пороговые значения								
γ_1	-8,487	1,246	-6,811	0,00	-8,487	-	-	-
γ_2	-4,045	0,496	-8,163	0,00	-4,316	0,207	-20,85	0,00
SNP-коэффициенты								
1	-	-	-	-	-1,870	0,579	-3,23	0,00
2	-	-	-	-	0,091	0,189	0,48	0,63
3	-	-	-	-	0,245	0,078	3,16	0,00

Таблица 3. Изменения в десятке наиболее рискованных для проживания населения городов/районов Оренбургской области

Ранг	1998 год		2005 год	
	Район/город	Значение ИП	Район/город	Значение ИП
1	г. Медногорск	-16,14	Бузулукский	-24,60
2	Бугурусланский	-12,82	Соль-Илецкий	-22,58
3	Соль-Илецкий	-9,24	г. Медногорск	-19,97
4	г. Орск	-8,91	г. Оренбург	-15,76
5	Пономаревский	-8,86	Абдулинский	-15,76
6	Кувандыкский	-7,10	г. Орск	-14,45
7	Сакмарский	-6,44	г. Новотроицк	-14,16
8	г. Новотроицк	-5,89	Грачевский	-13,42
9	Новоорский	-5,51	Пономаревский	-10,68
10	Бузулукский	-5,41	Тоцкий	-9,11

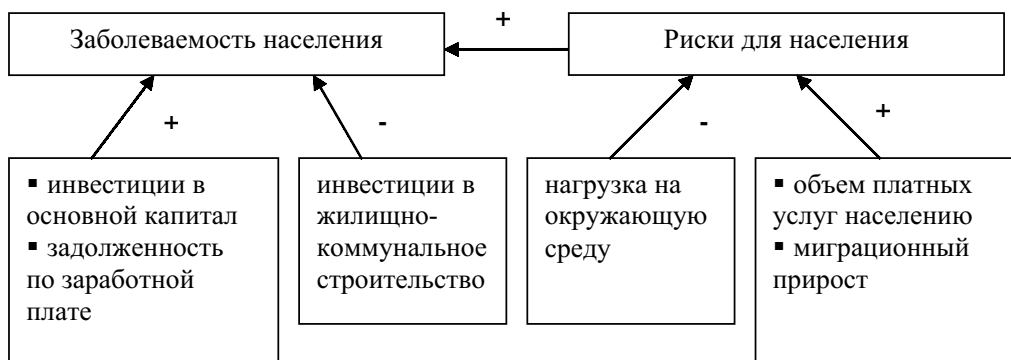


Рисунок 2. Схема предполагаемых взаимосвязей между переменными

органом государственной статистики по Оренбургской области).

Введем следующие обозначения:

RN – бинарная переменная, принимающая значение 0, если город или район характеризуется низким риском для населения от воздействия окружающей среды, значение 1 – высоким или средним риском;

UOZ – бинарная переменная, принимающая значение 0, если уровень общей заболеваемости в городе/районе ниже среднеобластного, значение 1 – выше (показатель превышения среднеобластной заболеваемости);

CER – бинарная переменная, характеризующая уровень социально-экономического развития города или района и принимающая значение 0, если город/район имеет низкий уровень развития, значение 1 – высокий уровень развития (по результатам многомерной классификации городов и районов области по показателям социально-экономического развития);

IPER – интегральный показатель, характеризующий нагрузку на окружающую среду (построен по аналогии с показателем риска для населения от воздействия окружающей среды);

MP – миграционный прирост (на 1 тыс. чел.);

VZD – ввод в действие жилых домов (на душу населения, руб.);

KB – количество безработных (на 1 тыс. чел., чел.);

OPUN – объем платных услуг населению (на душу населения, тыс. руб.);

IOK – инвестиции в основной капитал (на душу насел., тыс. руб.);

IZKS – инвестиции в жилищно-коммунальное строительство (на душу населения, тыс. руб.);

ZZP – задолженность предприятий и организаций по заработной плате (на 1 работника, тыс. руб.);

SZP – средняя заработная плата (тыс. руб.);

SOZ – средняя обеспеченность жильем (на душу населения, кв. м).

Предполагаемые взаимосвязи между переменными изображены на рисунке 2.

Указанные взаимосвязи описываются моделью вида:

$$\begin{cases} y_{RN,i}^* = \beta_{10} + \beta_{11} \cdot IPER_i + \beta_{12} \cdot OPUN_i + \\ + \beta_{13} \cdot MP_i + \epsilon_{RN,i} \\ y_{UOZ,i}^* = \beta_{20} + \beta_{21} \cdot RN_i + \beta_{22} \cdot IOK_i + \\ + \beta_{23} \cdot ZZP_i + \beta_{24} \cdot IZKS_i + \epsilon_{UOZ,i} \end{cases}, \quad (5)$$

где $i = 1, \dots, 312$.

В (5) $y_{RN,i}^*$ и $y_{UOZ,i}^*$ – латентные переменные, а RN и UOZ – такие бинарные переменные, что выполняется:

$$\begin{cases} RN_i = 1, y_{RN,i}^* > 0 \\ RN_i = 0, y_{RN,i}^* \leq 0 \end{cases}, \begin{cases} UOZ_i = 1, y_{UOZ,i}^* > 0 \\ UOZ_i = 0, y_{UOZ,i}^* \leq 0 \end{cases} \quad (6)$$

Закон совместного распределения $\epsilon_{RN,i}$ и $\epsilon_{UOZ,i}$ имеет вид:

$$\begin{pmatrix} \epsilon_{RN,i} \\ \epsilon_{UOZ,i} \end{pmatrix} \sim N(\mu, \Sigma), \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad (7)$$

где ρ – коэффициент ковариации ошибок $\epsilon_{RN,i}$ и $\epsilon_{UOZ,i}$.

Модель (5)–(7) представляет собой двумерную пробит-модель, во втором уравнении которой в качестве регрессора присутствует зависимая переменная первого урав-

нения. Это означает, что модель является рекурсивной системой одновременных уравнений, или так называемой двумерной пробит-моделью [4]. Система уравнений является идентифицируемой, оценивание модели методом максимального правдоподобия даст состоятельные и асимптотически эффективные оценки (оценки модели, полученные в пакете Stata, приведены в таблице 4).

Таким образом, снижение нагрузки на окружающую среду действительно будет вести к снижению рисков для населения от воздействия окружающей среды. Высокие риски для населения наблюдаются в городах и районах с высоким уровнем миграции и большим объемом платных услуг, оказываемых населению. В целом на риски для населения оказывают значительное влияние состояние окружающей среды и показатели, отражающие привлекательность региона.

Превышение среднеобластной заболеваемости связано с наличием высоких рисков для населения от воздействия окружающей среды, увеличением объема инвестиций в основной капитал и задолженностью по заработной плате перед населением. Это, возможно, объясняется тем, что инвестиции в основной капитал преимущественно направляются на расширение производственных возможностей предприятий, а не на техническое усовершенствование оборудования и ввод очистных сооружений с целью снижения негативного воздействия на окружающую среду.

Рассмотрим подробнее такие характеристики модели, как вероятность того, что в районе/городе одновременно и существует риск для населения, и наблюдается превышение среднеобластной заболеваемости $P(RN = 1, UOZ = 1)$, вероятность повышенной заболеваемости в городе/районе при условии существования в этом городе/районе

Таблица 4. Результаты оценивания двумерной модели, описывающей риски для населения и превышение среднеобластной заболеваемости населения

Показатель	Оценка коэффициента	Стандартное отклонение	z-статистика	Уровень значимости
<i>уравнение с зависимой переменной RN</i>				
IPER	-0,027	0,010	-2,67	0,008
OPUN	0,322	0,077	4,18	0,000
MP	0,023	0,011	2,09	0,037
cons	-0,656	0,141	-4,65	0,000
<i>уравнение с зависимой переменной UOZ</i>				
RN	1,028	0,306	3,36	0,001
IOK	0,037	0,016	2,38	0,018
ZZP	0,126	0,070	1,80	0,072
IZKS	-0,255	0,090	-2,84	0,004
cons	-0,513	0,201	-2,55	0,011
Wald chi2(7) = 71,42			Prob > chi2 = 0,0000	
chi2(1) = 7,0329			Prob > chi2 = 0,0080	

Таблица 5. Характеристики двумерной модели, описывающей риски для населения и вероятность превышения среднеобластной заболеваемости

Характеристика	Год								Среднее значение за 1998-2005 гг.
	1998	1999	2000	2001	2002	2003	2004	2005	
P (RN=1,UOZ=1)	0,14	0,14	0,16	0,21	0,20	0,27	0,27	0,30	0,21
P (RN=1,UOZ=0)	0,29	0,29	0,30	0,27	0,31	0,29	0,32	0,37	0,31
P (RN=0,UOZ=1)	0,36	0,36	0,35	0,34	0,33	0,32	0,27	0,22	0,32
P (RN=0,UOZ=0)	0,21	0,21	0,19	0,18	0,16	0,12	0,13	0,11	0,16
P (RN=1)	0,42	0,42	0,46	0,48	0,51	0,56	0,60	0,67	0,52
P (UOZ=1)	0,50	0,50	0,51	0,55	0,53	0,59	0,55	0,52	0,53
P (RN=1 UOZ=1)	0,22	0,22	0,26	0,30	0,32	0,40	0,41	0,48	0,33
P (UOZ=1 RN=1)	0,27	0,27	0,29	0,33	0,33	0,41	0,38	0,38	0,33

риска для населения ($P(UOZ = 1|RN = 1)$) и другие (таблица 5).

Из таблицы видно, что среднее за период с 1998 по 2005 г. значение вероятности того, что города и районы области одновременно характеризуются наличием риска для населения и повышенной по сравнению со среднеобластной заболеваемостью, увеличилось более чем в два раза (a в среднем составляет 0,21). Наибольшие вероятности возникновения такой ситуации наблюдаются в Бузулукском, Кувандыкском, Оренбургском и Новоорском районах, городах Медногорск, Новотроицк и Орск. Наиболее проблемными за последние годы стали Адамовский, Гайский, Светлинский, Октябрьский, Соль-Илецкий, Сорочинский, Тоцкий и Ясненский районы. Снижение вероятности одновременного существования риска для населения и превышения среднеобластной заболеваемости наблюдается в Новосергиевском, Переволоцком, Сакмарском и Первомайском районах.

В последнее время наиболее вероятной является ситуация, связанная с наличием значительного риска для населения от воздействия окружающей среды и отсутствием превышения средней заболеваемости населения. Такая ситуация в последние годы характерна для Бугурусланского, Саракташского, Северного, Ташлинского, Тюльганского районов и города Оренбурга. Учитывая, что предельная вероятность превышения среднеобластной заболеваемости для этих территорий значительно снижается с течением времени, а предельная ве-

роятность высокого риска для населения возрастает, можно предположить повышение в рассмотренных районах качества предоставляемых медицинских услуг и условий быта.

Средняя вероятность превышения среднеобластной заболеваемости в сочетании с низкими рисками для населения равна 0,32. При низких рисках для населения различия в уровнях заболеваемости для городов и районов области могут возникать, например, из-за наличия различных климатических, ландшафтных и других особенностей рассматриваемых территорий и требуют дальнейшего исследования.

Средняя вероятность того, что в городах и районах области отсутствуют и высокие риски для населения, и превышение среднеобластной заболеваемости, составляет всего 0,16. По отношению к 1998 году вероятность возникновения такой ситуации в 2005 году существенно уменьшилась, что свидетельствует об ухудшении общей ситуации в регионе.

Анализ вероятности наличия повышенной по сравнению со среднеобластной заболеваемости при условии высокого риска для населения показал значительное усиление этого влияния по сравнению с началом исследуемого периода. В 2005 году эта вероятность составила 0,48.

Анализ модели также показал, что вероятность повышенной заболеваемости в некотором городе/районе при наличии риска для населения от воздействия окружающей среды увеличивается на 0,39 по сравнению с ситуацией, когда риска для населения нет.

Список использованной литературы:

1. Руководство по оценке риска для здоровья населения при воздействии химических веществ, загрязняющих окружающую среду (Human Health Risk Assessment from Environmental Chemicals) – М.: Федеральный центр Госсанэпиднадзора Минздрава России, 2004. – 143 с.
2. Онищенко, Г.Г. Основы оценки риска для здоровья населения при воздействии химических веществ, загрязняющих окружающую среду / Под ред. Рахманина Ю.А., Онищенко Г.Г. – М.: НИИ ЭЧ и ГОС, 2002. – 408 с.
3. Носко В.П. Эконометрика для начинающих (дополнительные главы) / В.П. Носко. – М.: ИЭПП, 2005. – 379 с.
4. Тихомиров Н.П. Эконометрика / Н.П. Тихомиров, Е.Ю. Дорохина. – М.: Изд-во «Экзамен», 2003 – 512 с.
5. Johnson P.A. A test of the normality assumption in the ordered probit model // METRON – 1996, LIV, №3-4. – pages 213-221
6. Stewart, M. B. Semi-nonparametric estimation of extended ordered probit models // Stata Journal – 2004, vol. 4(1), pages 27-39