Сибирский экологический журнал, 2 (2013) 295–301

УДК 614.71:616.24-006-07(571.17)

Статистические методы исследования влияния выбросов загрязняющих веществ в атмосферу на заболеваемость населения Кемеровской области раком легкого

С. А. МУН, С. А. ЛАРИН, А. Н. ГЛУШКОВ

Институт экологии человека СО РАН
650065, Кемерово, просп. Ленинградский, 10
E-mail: Stellamun@yandex.ru

АННОТАЦИЯ

Описаны линейная, степенная, показательная и гиперболическая модели регрессионного анализа для выявления зависимости стандартизованных показателей заболеваемости раком легкого от количества выбросов загрязняющих веществ в атмосферу в Кемеровской области и представлены возможности их применения для расчета среднесрочных прогнозов канцерогенного воздействия окружающей среды на население Кузбасса.

Ключевые слова: модели регрессионного анализа, рак легкого, выбросы загрязняющих веществ в атмосферу, Кемеровская область.

Существенный спад производства в 90-е гг. прошлого столетия в Кемеровской области повлек за собой снижение уровня загрязнения атмосферного воздуха, что не могло не отразиться на показателях заболеваемости злокачественными опухолями, в том числе раком легкого (РЛ) [1].

Анализ динамики выбросов загрязняющих веществ (ЗВ) в атмосферу в Кемеровской области с 1985 по 2005 г. показал, что вплоть до 1997 г. их количество постоянно снижалось (с 2060,30 до 1196,16 тыс. т), затем наметилась тенденция к росту (в 2005 г. – 1676,33 тыс. т).

В 90-е гг. показатели заболеваемости РЛ в регионе колебались от 43 на 100 тыс. населения в 1992 г. до 51 в 1999 г. В промежутке с 2002 по 2010 г. заболеваемость РЛ оставалась примерно на одном уровне (39–43).

Очевидность описанных ранее взаимосвязей спада производства и соответствующего уменьшения количества выбросов ЗВ в атмосферу, с одной стороны, и снижения уровня заболеваемости РЛ – с другой, является основанием для более детального их исследования. При этом следует учитывать длительность латентного периода возникновения рака.

Для анализа искомых взаимосвязей и определения прогнозов заболеваемости используют как линейную, так и нелинейные (степенную, показательную, гиперболическую) математические модели [2, 3].

Задача настоящего исследования – построить адекватную математическую модель для расчета прогнозов заболеваемости РЛ населения Кемеровской области.

МАТЕРИАЛ И МЕТОДЫ ИССЛЕДОВАНИЯ

Данные о количестве выбросов ЗВ в атмосферу с 1985 по 2009 г. взяты из ежегодных Государственных докладов “О состоянии и охране окружающей среды Кемеровской об-

ласти" и "Угольная промышленность Российской Федерации" [4, 5].

Данные о количестве впервые выявленных случаев заболеваемости РЛ в Кемеровской области выбраны из основных форм медицинской документации ГУЗ "Областной клинический онкологический диспансер" г. Кемерово (форма № 7 "Сведения о заболеваемости ЗН") в промежутке с 1990 по 2010 г. Данные о возрастной структуре населения Кузбасса представлены Областным управлением статистики.

Информационную базу данных сформировали с помощью компьютерной программы "EXCEL-2000".

Расчет стандартизованных показателей заболеваемости РЛ (на 100 тыс. населения) проводили прямым методом стандартизации по общепринятой методике [6]. За стандарт принята возрастная структура населения Кемеровской области в 2001 г.

Математическую обработку результатов выполняли общепринятыми методами медицинской статистики с помощью компьютерной программы "EXCEL-2000" и пакета прикладных программ STATISTICA 6.0 (серийный номер № 31415926535897) [2, 3, 6].

Статистическая обработка информации строилась с учетом характера распределения данных (критерий Шапиро – Уилко, W).

Для выявления зависимости стандартизованных показателей РЛ (параметр y) с 1990 по 2005 г. от количества выбросов ЗВ в атмосферу (фактор x) с 1985 г. по 2005 г. в Кемеровском регионе использовали метод расчета коэффициента прямой, линейной (парной) корреляции с выявлением статистически значимого коэффициента корреляции ($p < 0{,}05$) и определением временного сдвига (t) между параметром y и фактором x. При построении уравнения регрессии (однофакторный регрессионный анализ) проверяли адекватность моделей для определения влияния фактора x на параметр y.

Алгоритм линейной и нелинейных моделей регрессионного анализа включал в себя: расчет параметров уравнения регрессии и коэффициентов регрессии (a и b) с проверкой значимости (t-критерий Стьюдента, $p = 0{,}05$); вычисление остаточного компонента (R/S-критерий); проверку выполнения предпосылок метода наименьших квадратов (МНК); расчет коэффициента детерминации с проверкой значимости (F-критерий Фишера, $p = 0{,}05$), стандартной ошибки модели (S_e), коэффициента эластичности ($Э_i$) и средней относительной ошибки аппроксимации ($E_{отн\_i}$).

После оценки моделей рассчитали прогноз показателей заболеваемости РЛ на 2006–2017 гг. по имеющимся данным о количестве выбросов ЗВ в атмосферу в предыдущие годы (1998–2010). Полученные таким образом ожидаемые показатели заболеваемости населения РЛ с 2006 по 2010 г. сопоставлены с фактическими в этот период.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Установлено, что показатели заболеваемости РЛ населения Кемеровской области и количества выбросов ЗВ в атмосферу имеют нормальное распределение по критерию Шапиро – Уилко ($W = 0{,}929$; $p = 0{,}147$ и $W = 0{,}959$; $p = 0{,}400$ соответственно).

На рис. 1, *А* показана взаимосвязь во времени показателей заболеваемости РЛ и количества выбросов ЗВ в атмосферу с промежутком t между ними, равным 8 лет. По рис. 1, *В* видно, что между ними существует прямая, сильная корреляционная связь ($r = 0{,}79$; $p = 0{,}001$).

Сначала мы рассмотрели **линейную модель**: $\hat{y} = a + b \times x$.

Параметры уравнения линейной регрессии приведены в табл. 1.

Уравнение регрессии имеет следующий вид: $\hat{y} = 25{,}532 + 0{,}013 \times x$.

Коэффициенты регрессии $a = 25{,}53$ ($t = 5{,}211$; $p = 0{,}0003$), $b = 0{,}013$ ($t = 4{,}241$; $p = 0{,}0014$). Коэффициент b означает, что при увеличении количества выбросов ЗВ в атмосферу на 1 тыс. т в год заболеваемость РЛ увеличивается в среднем на 0,013 на 100 тыс. населения.

Выполнение предпосылок МНК согласно условиям Гаусса – Маркова включают в себя проверку: 1) случайности остаточной компоненты (критерий поворотных точек); 2) равенства нулю математического ожидания остаточной компоненты $\bar{E} = 0$ и постоянства дисперсии (критерий Голдфелда – Квандта); 3) независимости уровней ряда остатков (кри-

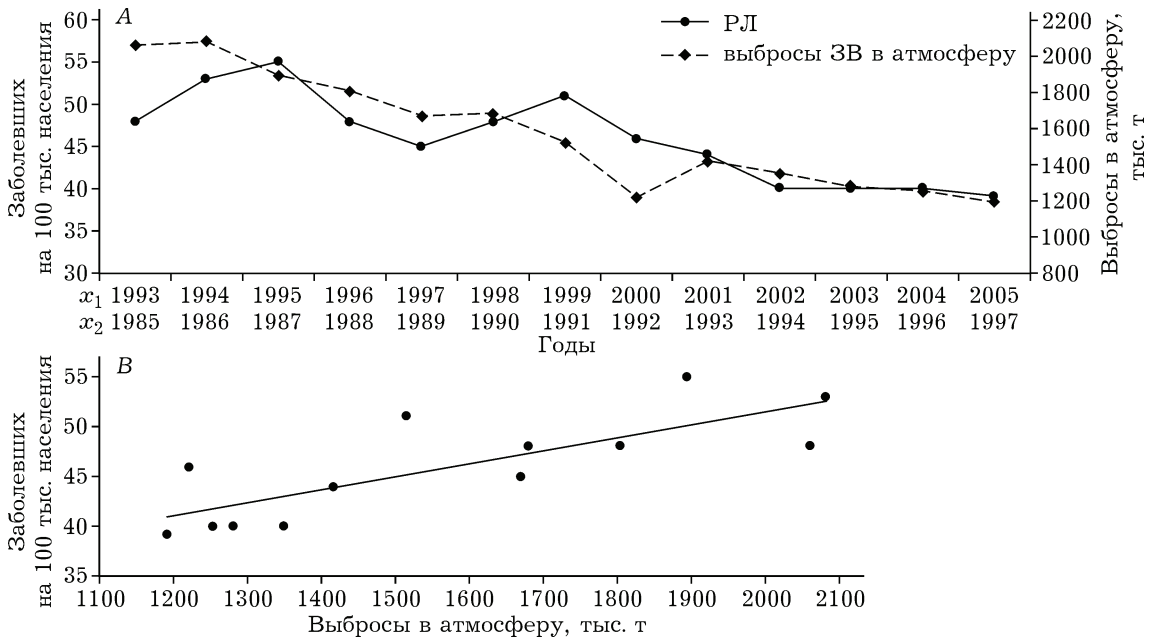

Рис. 1. Взаимосвязь показателей заболеваемости населения раком легкого (РЛ) с выбросами загрязняющих веществ (ЗВ) в атмосферу в Кемеровской области.

A – во времени: x_1 – интервал заболеваемости раком легкого с 1993 по 2005 г., x_2 – интервал выбросов загрязняющих веществ в атмосферу с 1985 по 1997 г., *B* – корреляция

терий Дарбина – Уотсона); 4) соответствия ряда остатков закону распределения (R/S-критерий).

1. По графику остатков (рис. 2) определяли количество поворотных точек. В нашем случае их 8 ($p = 8$) при $p_{кр} = [2,5] = 2$. Критическое значение вычисляли по формуле

$$p_{кр} = \left[\frac{2(n-2)}{3} - 1,96\sqrt{\frac{16n-29}{30}}\right]$$

при $n = 13$ (количество наблюдений).

Следовательно, свойство случайности для ряда остатков выполняется: $p = 8 > p_{кр} = 2$.

2. В данной модели выполняется и второе условие Гаусса–Маркова: равенство нулю математического ожидания $\bar{E} = -2,186231E - 15 = 0$.

При проверке постоянства дисперсии на гетероскедастичность (критерий Голдфелда–Квандта) $F = FS_2/FS_1 = 39,42/29,68 = 1,33$ (где FS_1 и FS_2 – остаточная сумма квадратов по первым и последним пяти наблюдениям нашей модели). Критическое значение при уровне $\alpha = 5$ % и числах степеней свобода $k_1 = k_2 = 5-1-1=3$ составило $F_{кр} = 9,28$. Следовательно, наша линейная модель соответствует неравенству $F_{кр} = 9,28 > F = 1,33$. Это говорит о постоянстве дисперсии остатков, т. е. модель гомоскедастичная.

Таблица 1

Параметры уравнений линейной и нелинейной регрессий между показателями заболеваемости населения раком легкого и количеством выбросов загрязняющих веществ в Кемеровской области

Модель	*a*	*p*-значение	*b*	*p*-значение	R^2	S_e	$\bar{Э}_i$, %	$E_{отн\_i}$, %
Линейная	25,5324	0,0003	0,0130	0,0014	0,62	3,362	0,44	5,6
Степенная	0,2075	0,5414*	0,4554	0,0010	0,64	0,031	0,46	5,3
Показательная	1,4662	3,438E-12	0,000123	0,0013	0,62	0,032	0,45	5,4
Гиперболическая	67,3361	3,264E-08	−32425,7293	0,0011	0,64	3,291	0,44	5,4

* $p > 0,05$.

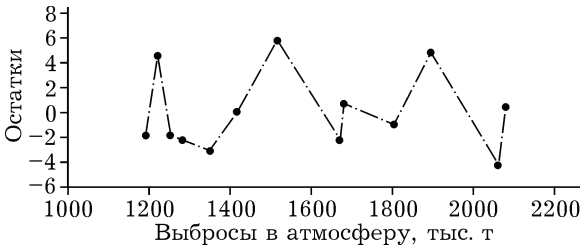

Рис. 2. График остатков линейной регрессии

3. Для проверки независимости уровней ряда остатков использовали критерий d (Дарбина – Уотсона, STATISTICA 6.0). Критерий d = 1,81 и эта величина больше d_U = 1,34 и меньше $4-d_U$ = 2,66, следовательно, автокорреляция отсутствует (H_0 : ρ = 0).

4. Проверка соответствия ряда остатков закону распределения (R/S-критерий) показала, что для построенной модели свойство нормального распределения остаточной компоненты выполняется согласно формуле

$$R / S = \frac{e_{\max} - e_{\min}}{S_e} = 2,99 ,$$

где $e_{\max}$ и $e_{\min}$ – значения остатков, S_e – стандартная ошибка оценки ($e_{\max}$ = 5,79; $e_{\min}$ = = – 4,28; S_e = 3,36).

Критический интервал границ $R/S_{\text{кр}}$ = = 2,86–4,0 при n = 13. Следовательно, для построенной линейной модели свойство нормального распределения остаточной компоненты выполняется (2,86 < 2,99 < 4,0).

Таким образом, проведенная проверка предпосылок регрессионного анализа показала, что для линейной модели выполняются все условия Гаусса – Маркова.

Коэффициент детерминации и статистическую значимость по F-критерию Фишера вычисляли по формулам (1) и (2):

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \overline{y})^2}, \qquad (1)$$

$$F = \frac{R^2}{1 - R^2} \times (n - 2). \qquad (2)$$

Значение R^2 = 0,62 при $F(1,11)$ = 17,983; p = 0,0014.

Остаточная сумма квадратов $SS_{\text{ост}}$ и дисперсия остатков MS составили 124,316 и 11,301 соответственно (табл. 2). Следовательно, влияние выбросов ЗВ в атмосферу на заболеваемость населения РЛ составляет 62 %, остальные 38 % следует отнести к неучтенным факторам.

Стандартная ошибка регрессии

$$S_e = \sqrt{\frac{SS_{\text{ост}}}{n - 2}} = 3,362. \qquad (3)$$

Это говорит о том, что, чем меньше в целом разброс точек наблюдений около прямой регрессии, тем надежней будет уравнение как оценочная функция.

Полученный по формуле коэффициент эластичности $Э_i = b \times \frac{\overline{x}_i}{\overline{y}_i} = 0,44\ \%$. Это значит, что при увеличении выбросов ЗВ в атмосферу на 1 % заболеваемость населения РЛ возрастает на 0,44 %.

Проверка значимости полученного уравнения регрессии с помощью F-критерия Фишера показала, что уравнение модели является значимым при F = 17,983 > $F_{\text{кр}}$= 4,84 (p = 0,0014) и его можно использовать для дальнейшего анализа, т. е. зависимая переменная $\hat{y}$ хорошо описывается включенной в модель факторной переменной x.

Для определения точности выбранной модели вычисляли среднюю относительную ошибку аппроксимации по формуле

$$E_{отн\_i} = \frac{1}{n} \times \left| \frac{\overline{e}_i}{\overline{y}_i} \right| \times 100\ \%, \qquad (4)$$

где $\left| \frac{\overline{e}_i}{\overline{y}_i} \right| \times 100\ \%$ – среднее значение относительной погрешности остатков.

Т а б л и ц а 2

Дисперсионный анализ показателей количества выбросов ЗВ в атмосферу и заболеваемости населения Кемеровской области раком легкого

Параметр	*df*	Сумма квадратов, *SS*	Дисперсия, *MS*	*F*-критерий	$F_{кр}$	*p*-значение	Степень влияния, %
Регрессия	1	203,232	203,232	17,983	4,84	0,0014	62
Остаток	11	124,316	11,301				38
Итого	12	327,548					100

Значение $E_{отн\_i}$ =5,6 % < 10 %, следовательно, модель точна.

Таким образом, на основании проверки предпосылок МНК, значений t-критерия Стьюдента, F-критерия Фишера и величины R^2 линейную модель можно считать адекватной для описания взаимосвязи количества выбросов ЗВ в атмосферу и заболеваемости населения РЛ Кемеровской области, а также использовать ее для прогнозирования заболеваемости РЛ в реальных условиях.

Далее мы рассмотрели нелинейные модели регрессионного анализа (см. табл. 1).

Степенная модель: $\hat{y} = a \times x^b$.

Исходную модель путем логарифмирования $\lg \hat{y} = \lg a + b \times \lg x$ преобразовали в линейную модель. Уравнение регрессии приобрело следующий вид: $\hat{y} = a + b \times x$.

Коэффициенты b и a рассчитали по формулам

$$b = \frac{\overline{xy} - \bar{y} \times \bar{x}}{\overline{x^2} - \bar{x}^2} = 0,4554; \quad t = 4,413; \; p = 0,001;$$

$$a = \bar{y} - b \times \bar{x} = 0,2075; \; t = 0,630; \; p = 0,541.$$

Выполнив потенцирование линейного уравнения, мы получили следующее уравнение:

$$\hat{y} = 10^{0,2075} \times x^{0,4554} = 1,613 \times x^{0,4554}.$$

Вычислили по формулам:

коэффициент детерминации R^2 = 0,64 (1) при $F(1,11)$ = 19,474, p = 0,001 (2);

стандартную ошибку регрессии (3) S_e = = 0,031;

коэффициент эластичности

$$Э_i = \frac{a \times b \times (\bar{x})^{b-1} \times \bar{x}}{a \times (\bar{x})^b} = 0,46 \%;$$

среднюю ошибку аппроксимации (4) $E_{отн\_i}$ = = 5,3 %.

Таким образом, в степенной модели коэффициент ***a*** оказался статистически незначимым ($p > 0,05$), следовательно, использование этой модели для прогнозирования нецелесообразно.

Показательная модель: $\hat{y} = a \times b^x$.

Для построения этой модели также произвели линеаризацию переменных путем логарифмирования обеих частей уравнения $\lg \hat{y} = \lg a + x \times \lg b$ и получили следующее линейное уравнение регрессии: $\hat{y} = a + b \times x$.

Коэффициенты b и a рассчитаны по формулам

$$b = \frac{\lg \overline{y \times x} - \lg \bar{y} \times \bar{x}}{\overline{x^2} - \bar{x}^2} = 0,000123; \; t = 4,281;$$

$$p = 0,001;$$

$$a = \bar{y} - b \times \bar{x} = 1,466; \; t = 31,882; \; p = 3,438\text{E-}12.$$

После потенцирования линейного уравнения регрессии получили следующее уравнение:

$$\hat{y} = 10^{1,466} \times (10^{0,000123})^x = 29,2415 \times 1,000283^x.$$

Далее рассчитали: R^2 = 0,62 (1) при $F(1,11)$ = = 17,983, p = 0,0014 (2), S_e = 0,032 (3), $E_{отн\_i}$= = 5,4 % (4) и

$$Э_i = \frac{a \times b^{\bar{x}} \times \ln b \times \bar{x}}{a \times b^{\bar{x}}} = 0,45 \%.$$

Таким образом, параметры показателей степенной модели оказались статистически значимыми, поэтому модель может быть использована для расчета прогнозов заболеваемости РЛ.

Гиперболическая модель: $\hat{y} = a + b/x$.

Линеаризацию модели произвели путем замены ***x*** на $1/x = X$ и получили следующее линейное уравнение: $\hat{y} = a + b \times X$.

Коэффициенты регрессии:

$$b = \frac{\overline{y \times X} - \bar{y} \times \bar{X}}{\overline{X^2} - \bar{X}^2} = -32425,729;$$

$$a = \bar{y} - b \times \bar{X} = 67,336;$$

$t = -4,387$, p = 0,001 и t = 13,567, p = 3,264E-08 соответственно.

Значения R^2 = 0,64 (1) при $F(1,11)$ = 19,251, p = 0,0011 (2), S_e = 3,291 (3),

$E_{отн\_i}$=5,4 % (4) и

$$Э_i = b \times \frac{1}{\bar{x}^2} \times \frac{\bar{x}}{a - b/\bar{x}} = \frac{b}{a \times \bar{x} - b} = 0,44 \%.$$

Таким образом, параметры показателей гиперболической модели оказались статистически значимыми, следовательно, модель может быть использована для расчета прогнозов заболеваемости РЛ.

По линейной, показательной и гиперболической моделям мы рассчитали прогнозируемые средние значения показателя заболеваемости населения РЛ ($\hat{\boldsymbol{y}}$) на 2006–2017 гг. при уровне значимости p = 0,05 и 80 % от имеющихся фактических данных количества выбросов ЗВ в атмосферу (x^*) за 1988–2010 гг.

Таблица 3

Сопоставление фактических (y) и расчетных (прогнозируемых, $\hat{y}$) показателей заболеваемости населения Кемеровской области раком легкого в 2006–2010 гг.

Год	1998	1999	2000	2001	2002
Фактические показатели выбросов ЗВ, тыс. т ($\boldsymbol{x}$)	990,124	1086,8584	1184,754	1292,263	1236,925
Год	2006	2007	2008	2009	2010
Фактические показатели заболеваемости РЛ ($\boldsymbol{y}$)	39	39	38	43	40
Прогнозируемые показатели заболеваемости РЛ ($\boldsymbol{y}$)					
Линейная модель	38,4 (1,6 %)	39,6 (1,6 %)	40,9 (7,7 %)	42,3 (1,6 %)	41,6 (4,0 %)
Доверительный интервал, p = 0,05	34,0–42,8	35,8–43,5	37,6–44,2	39,6–45,1	38,6–44,6
Показательная модель	**38,7 (0,8 %)**	**39,8 (2,0 %)**	**40,9 (7,6 %)**	**42,2 (2,0 %)**	**41,5 (3,7 %)**
Доверительный интервал, p = 0,05	**38,6–38,8**	**39,7–39,8**	**40,8–40,9**	**42,1–42,2**	**41,4–41,5**
Гиперболическая модель	34,6 (11,3 %)	37,5 (3,8 %)	40,0 (5,2 %)	42,2 (1,8 %)	41,4 (2,8 %)
Доверительный интервал, p = 0,05	30,3–38,9	33,8–41,2	36,8–43,2	39,5–44,9	38,2–44,1

Примечание. В скобках указаны отклонения прогнозируемых показателей от фактических.

Расчетные прогнозируемые показатели заболеваемости населения РЛ ($\hat{\boldsymbol{y}}$) за 2006–2010 гг. по описанным моделям сопоставили с фактическими за тот же период времени ($\boldsymbol{y}$). Полученные результаты представлены в табл. 3.

В результате выяснилось, что прогнозируемые показатели заболеваемости РЛ отклоняются от фактических меньше всего в показательной модели – на 0,8–7,6 %, в линейной модели – на 1,6–7,7, в гиперболической – на 1,8–11,3 %.

Точность и надежность прогнозируемых значений $\hat{\boldsymbol{y}}$ в каждой модели оценивали 95%-м доверительным интервалом по формуле:

$$Y = \hat{y} \pm t_{95} \times m_y,$$

$$m_{\hat{y}} = S_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}},$$

где $\hat{\boldsymbol{y}}$ – прогнозируемые значения показателя заболеваемости населения РЛ, $t_{95}(11) = 2{,}2$, S_e – стандартная ошибка модели (для линейной – 3,362; показательной – 0,032, гиперболической – 3,291); $m_{\hat{y}}$ – стандартная ошибка прогнозирования.

На рис. 3 представлен прогноз заболеваемости РЛ населения Кемеровской области с 2011 по 2017 г. с помощью показательной модели на основе фактических данных о количестве выбросов ЗВ в атмосферу за 2003–2010 гг. С 2011 по 2016 г. заболеваемость РЛ населения Кемеровской области будет расти до 44,1 на 100 тыс. населения, а к 2017 г. показатель заболеваемости РЛ составит 43,3.

Таким образом, полученные в линейной, показательной и гиперболической моделях коэффициенты регрессий $\boldsymbol{a}$ и $\boldsymbol{b}$ статистически значимы ($p < 0{,}05$), имеют высокие коэффициенты детерминации R^2 = 62–64, связь между фактором $\boldsymbol{x}$ и результативным признаком $\hat{\boldsymbol{y}}$ у моделей одинаковая $Э_i$ = 0,44–0,45 %. Наиболее точной оказались показательная и гиперболическая модели с ошибкой аппроксимации $E_{отн\_i}$= 5,4 %, т. е. в среднем расчетные значения $\hat{\boldsymbol{y}}$ для этих моделей отличались от фактических значений на 5,4 %, но у показательной модели стандартная ошибка регрессии оказалась самой низкой S_e = 0,031 (см. табл. 1).

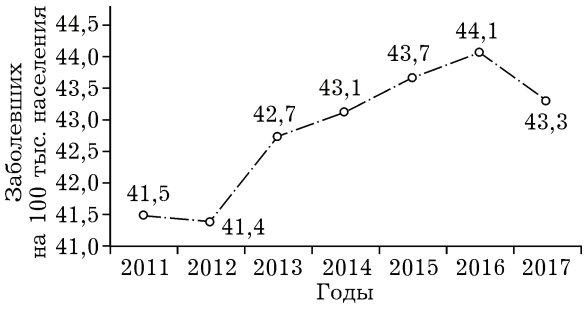

Рис. 3. Прогноз заболеваемости раком легкого населения Кемеровской области на 2011–2017 гг.

ЗАКЛЮЧЕНИЕ

При анализе взаимосвязей показателей заболеваемости населения раком легкого с количеством выбросов загрязняющих веществ в атмосферу в Кемеровской области показательная модель оказалась наиболее адекватной по всем статистическим условиям и критериям. Рассчитанные с помощью этой модели показатели заболеваемости раком легкого за 2006–2010 гг. отличаются от фактических не более чем на 7,6 % с надежностью 95 %. Поэтому показательная модель и была использована для дальнейшего прогнозирования заболеваемости раком легкого населения Кемеровской области с 2011 по 2017 г. с учетом фактических данных выбросов загрязняющих веществ в атмосферу с 2003 по 2010 г. По нашим расчетам ожидается рост показателя заболеваемости раком легкого с 40,0 на 100 тыс. населения в 2010 г. до 43,3–44,1 в 2016–2017 гг.

ЛИТЕРАТУРА

1. Мун С. А., Ларин С. А., Глушков А. Н., Счастливцев Е. Л., Браиловский В. В., Чухров Ю. С., Магарилл Ю. А. Ретроспективный анализ и прогноз заболеваемости раком легкого в Кузбассе // Медицина труда и промышленная экология. 2007. № 12. С. 22–26.
2. Степанов В. Г. Эконометрика. URL: http://www.e-college.ru/xbooks/xbook019/book/index/index.html?go=part-011*page.htm
3. Халафян А. А. STATISTICA 6. Статистический анализ данных. М.: ООО "Бином-Пресс", 2008.
4. Государственный доклад "О состоянии и охране окружающей среды Кемеровской области". URL: http://www.ecokem.ru/00004.html
5. Угольная промышленность Российской Федерации в 2004 году. Т. II. М.: Росинформуголь, 2005.
6. Мерков А. М., Поляков Л. Е. Санитарная статистика. Л.: Медицина, 1974.

Statistical Methods for the Investigation of the Effects of Pollutant Emissions into the Atmosphere on Lung Cancer Morbidity of the Population of Kemerovo Region

S. A. MUN, S. A. LARIN, A. N. GLUSHKOV

Institute of Human Ecology SB RAS
650065, Kemerovo, Leningradskiy ave., 10
E-mail: Stellamun@yandex.ru

Linear, power, exponential and hyperbolic models of regression analysis for the determination of the dependence of standardized indices of lung cancer morbidity on the amount of pollutants emitted into the atmosphere in the Kemerovo Region are described. The possibilities of the application of these models to the calculation of intermediate-term forecasts of carcinogenic action of the environment on the population of the Kuznetsk Basin are discussed.

Key words: models of regression analysis, lung cancer, emission of pollutants into the atmosphere, Kemerovo Region.