

Информационные технологии и управление в области безопасности жизнедеятельности

УДК 311+519.23/.25

Регрессионный анализ временного ряда количества пожаров в России

Regressionnyy analysis of the temporary row amount in Russia

*И.А. Кайбичев¹,
д-р физ-мат. наук, доцент;
Е.И. Кайбичева²,
канд. эконом. наук*

¹ФГБОУ ВО Уральский институт ГПС МЧС России;

²Уральский государственный экономический университет

*I.A. Kaibichev¹,
Doctor in Physical and
Mathematic Sciences;*

*E.I. Kaibicheva²,
Ph.D. of Economic Sciences*

¹The Ural Institute of State
Firefighting Service of Ministry
of Russian Federation for Civil
Defense;

²The Ural State University of
Economics

Аннотация:

Рассмотрены возможности аппроксимации временного ряда количества пожаров в России за 2001-2017 года наиболее часто используемыми математическими функциями.

Ключевые слова: математическое моделирование, временные ряды, число пожаров, Российская Федерация.

Abstract:

The considered possibilities to approximations of the temporary row amount fire in Russia for 2001-2017 most often used mathematical function.

Key words: mathematical modeling, temporary rows, number fire, Russian Federation.

Введение

Необходимость построения математической модели количества пожаров в России возникла в связи с потребностью прогнозирования этого показателя на следующий год с целью обоснования планирования обеспечения ГПС МЧС России.

Имеющиеся данные по количеству пожаров в России за 2001-2017 года [1-13] представляют собой временной ряд [14-16]. Такой ряд состоит из числовых значений количества пожаров и периода времени, по состоянию на которое даны эти числовые значения (год). Временной ряд отличается от простой выборки данных тем, что при анализе необходимо учитывать взаимосвязь измерений со временем, а не только характеристики выборки.

Математический аппарат теории временных рядов, в частности автокорреляционные функции, применялся для построения математических моделей количества пожаров в Красноярском и Забайкальском краях, Новосибирской и Иркутской областях [17,18].

Цель данного исследования – выполнить корреляционный и регрессионный анализы данных о пожарах в России за 2001-2017 года, рассмотреть возможности их аппроксимации часто используемыми функциями, обосновать выбор наиболее подходящей функции.

Корреляционный анализ

Ранее было установлено наличие временной зависимости числа пожаров в России [19]. Фактор времени учитывался путем задания номера года.

Отметим, что в работе [19] рассматривались данные 2001-2015 годов. Поэтому рассмотрим статистику за 2001-2017 года (Таб. 1).

Таблица 1. Количество пожаров в Российской Федерации (тыс. ед) за 2001-2017 года

Год	Пожары
2001	246,5
2002	260,8
2003	239,2
2004	233,2
2005	229,8
2006	220,5
2007	212,6
2008	202,2
2009	187,6
2010	179,5
2011	168,5
2012	162,9
2013	153,5
2014	150,8
2015	145,9
2016	139,5
2017	133,1

Рассчитаем коэффициент линейной корреляции Пирсона R [20] между количеством пожаров и номером года. В нашем случае R = - 0,99. Отметим, что для расчета была применена функция Коррел программы V=Microsoft Excel.

Имеем случай малой (n = 17 < 100), тогда выполняют перерасчет по формуле

$$R' = R \left[1 + \frac{1-R^2}{2(n-3)} \right] \quad (1)$$

Для нашего случая R' = - 0,99. Полученное значение коэффициента линейной корреляции Пирсона по модулю близко к 1. Это доказывает факт существования очень сильной и близкой к линейной связи между количеством пожаров и номером года. Отметим, что связь является отрицательной. При увеличении номера года число пожаров уменьшается.

Регрессионный анализ

Регрессионный анализ устанавливает параметры зависимости между исследуемыми факторами. Ранее [21] связь между числом пожаров в 2001-2015 годах и номером года аппроксимировали линейной функцией

$$Y_m = -8,4 * X + 17137 \quad (2)$$

где Y_м – модельное количество пожаров (тыс. ед.), X – номер года.

Реальные значения переменной Y как правило отличаются от модельных величин Y_м, рассчитанных по функции регрессии. Ошибку определяют как разность этих величин

$$\epsilon_i = Y_i - Y_{mi} \quad (3)$$

где ε_i – ошибка для i-го наблюдения, Y_i - фактическое значение, Y_{mi} – модельное значение, определенное по функции регрессии. Величину ошибки не удается объяснить влиянием факторов. Часто считают её случайной величиной.

Поэтому в качестве показателя качества регрессионной модели часто используют среднее значение

$$\bar{\epsilon^2} = \frac{1}{n} \sum_{i=1}^n \epsilon_i^2 \quad (4)$$

Выполним регрессионный анализ данных по числу пожаров в России за 2001-2017 года.

Первоначально исследуем возможность аппроксимации линейной функцией. Использование метода наименьших квадратов привело к результату

$$Y_m = -8,12985 * X + 16525 \quad (5)$$

Среднее значение квадрата ошибки составило 31,5 (Табл. 2).

Таблица 2. Линейная модель

Год	Y	Y _м	ε	ε ²
2001	246,5	257,2	-10,7	113,7
2002	260,8	249,0	11,8	138,5
2003	239,2	240,9	-1,7	2,9
2004	233,2	232,8	0,4	0,2
2005	229,8	224,6	5,2	26,6
2006	220,5	216,5	4,0	15,9
2007	212,6	208,4	4,2	17,8
2008	202,2	200,3	1,9	3,8
2009	187,6	192,1	-4,5	20,5
2010	179,5	184,0	-4,5	20,2
2011	168,5	175,9	-7,4	54,2
2012	162,9	167,7	-4,8	23,4
2013	153,5	159,6	-6,1	37,3
2014	150,8	151,5	-0,7	0,5
2015	145,9	143,3	2,6	6,5
2016	139,5	135,2	4,3	18,4
2017	133,1	127,1	6,0	36,2
среднее	192,1	192,1	0,0	31,5

Отметим, что среднее значение ошибки равно 0. Коэффициент линейной корреляции Пирсона между ошибкой и номером года также равен 0. Это позволяет считать ошибку случайной величиной.

Далее изучим возможность аппроксимации гиперболической функцией. Метод наименьших квадратов дал результат.

$$Y_m = 122,226/(X-2000)+167,3939 \quad (6)$$

Среднее значение квадрата ошибки составило 834,1 (Табл. 3).

Поэтому гиперболическая функция (6) дает худшие результаты по сравнению с линейной функцией (5). Коэффициент линейной корреляции Пирсона между ошибкой и номером года составил – 0,67. Следовательно, между ошибкой и номером года существует средняя корреляционная связь. Из-за большой величины среднего квадрата ошибки гиперболическая модель не годится для аппроксимации количества пожаров.

Таблица 3. Гиперболическая модель

Год	Y	Y _m	ε	ε ²
2001	246,5	289,6	-43,1	1859,3
2002	260,8	228,5	32,3	1042,8
2003	239,2	208,1	31,1	965,0
2004	233,2	198,0	35,2	1242,5
2005	229,8	191,8	38,0	1441,0
2006	220,5	187,8	32,7	1071,6
2007	212,6	184,9	27,7	769,8
2008	202,2	182,7	19,5	381,3
2009	187,6	181,0	6,6	43,9
2010	179,5	179,6	-0,1	0,0
2011	168,5	178,5	-10,0	100,1
2012	162,9	177,6	-14,7	215,5
2013	153,5	176,8	-23,3	542,7
2014	150,8	176,1	-25,3	641,3
2015	145,9	175,5	-29,6	878,7
2016	139,5	175,0	-35,5	1262,6
2017	133,1	174,6	-41,5	1720,9
среднее	192,1	192,1	0,0	834,1

Рассмотрим показательную функцию. Метод наименьших квадратов дал результат

$$Y_m = 274,925 * 0,95873^{X-2000} \quad (7)$$

Среднее значение квадрата ошибки составило 37 (Табл. 4). Эта величина больше чем в линейной модели (5). Коэффициент линейной корреляции Пирсона между ошибкой и номером года составил -0,08. Поэтому ошибку можно считать случайной величиной.

Коэффициент линейной корреляции Пирсона между ошибкой и номером года равен 0. Поэтому ошибку можно считать случайной величиной.

Попробуем степенную функцию. Метод наименьших квадратов дал результат

$$Y_m = 286,8336 * (X-2000)^{-0,20965} \quad (9)$$

Среднее значение квадрата ошибки составило 358,2 (Табл. 6). Это значение больше чем в линейной (5) и параболической (8) моделях. Коэффици-

ент линейной корреляции Пирсона между ошибкой и номером года равен -0,49.

Таблица 4. Показательная модель

Год	Y	Y _m	ε	ε ²
2001	246,5	263,6	-17,1	291,7
2002	260,8	252,7	8,1	65,6
2003	239,2	242,3	-3,1	9,4
2004	233,2	232,3	0,9	0,9
2005	229,8	222,7	7,1	50,6
2006	220,5	213,5	7,0	49,0
2007	212,6	204,7	7,9	62,6
2008	202,2	196,2	6,0	35,5
2009	187,6	188,1	-0,5	0,3
2010	179,5	180,4	-0,9	0,8
2011	168,5	172,9	-4,4	19,6
2012	162,9	165,8	-2,9	8,4
2013	153,5	159,0	-5,5	29,7
2014	150,8	152,4	-1,6	2,5
2015	145,9	146,1	-0,2	0,0
2016	139,5	140,1	-0,6	0,3
2017	133,1	134,3	-1,2	1,4
среднее	192,1	192,2	-0,1	37,0

Следовательно, между ошибкой и номером года существует средняя корреляционная связь. Из-за большой величины среднего квадрата ошибка степенная модель не подходит для аппроксимации числа пожаров.

Таблица 5. Параболическая модель

Год	Y	Y _m	ε	ε ²
2001	246,5	259,6	-13,1	172,1
2002	260,8	250,6	10,2	104,7
2003	239,2	241,6	-2,4	6,0
2004	233,2	232,8	0,4	0,1
2005	229,8	224,2	5,6	31,9
2006	220,5	215,6	4,9	24,1
2007	212,6	207,2	5,4	29,6
2008	202,2	198,8	3,4	11,3
2009	187,6	190,7	-3,1	9,3
2010	179,5	182,6	-3,1	9,5
2011	168,5	174,6	-6,1	37,7
2012	162,9	166,8	-3,9	15,3
2013	153,5	159,1	-5,6	31,5
2014	150,8	151,5	-0,7	0,5
2015	145,9	144,1	1,8	3,3
2016	139,5	136,7	2,8	7,6
2017	133,1	129,5	3,6	12,7
среднее	192,1	192,1	0,0	29,8

Исследуем возможность аппроксимации логарифмической функцией. Метод наименьших квадратов дал результат

$$Y_m = 286,6929 - 47,9833 * \ln(X - 2000) \quad (10)$$

Среднее значение квадрата ошибки составило 245,6 (Табл. 7). Это значение больше чем в линейной (5) и параболической (8) моделях.

Таблица 6. Степенная модель

Год	Y	Y _m	ε	ε ²
2001	246,5	286,8	-40,3	1626,8
2002	260,8	248,0	12,8	162,9
2003	239,2	227,8	11,4	129,4
2004	233,2	214,5	18,7	350,1
2005	229,8	204,7	25,1	630,7
2006	220,5	197,0	23,5	551,8
2007	212,6	190,7	21,9	477,7
2008	202,2	185,5	16,7	279,6
2009	187,6	181,0	6,6	44,2
2010	179,5	177,0	2,5	6,2
2011	168,5	173,5	-5,0	25,0
2012	162,9	170,4	-7,5	55,7
2013	153,5	167,5	-14,0	196,8
2014	150,8	164,9	-14,1	200,1
2015	145,9	162,6	-16,7	278,1
2016	139,5	160,4	-20,9	436,4
2017	133,1	158,4	-25,3	638,4
среднее	192,1	192,4	-0,3	358,2

Таблица 7. Логарифмическая модель

Год	Y	Y _m	ε	ε ²
2001	246,5	286,7	-40,2	1615,5
2002	260,8	253,4	7,4	54,3
2003	239,2	234,0	5,2	27,3
2004	233,2	220,2	13,0	169,7
2005	229,8	209,5	20,3	413,4
2006	220,5	200,7	19,8	391,3
2007	212,6	193,3	19,3	371,6
2008	202,2	186,9	15,3	233,6
2009	187,6	181,3	6,3	40,2
2010	179,5	176,2	3,3	10,8
2011	168,5	171,6	-3,1	9,8
2012	162,9	167,5	-4,6	20,8
2013	153,5	163,6	-10,1	102,4
2014	150,8	160,1	-9,3	85,8
2015	145,9	156,8	-10,9	117,8
2016	139,5	153,7	-14,2	200,4
2017	133,1	150,7	-17,6	311,4
среднее	192,1	192,1	0,0	245,6

Коэффициент линейной корреляции Пирсона между ошибкой и номером года равен -0,33. Поэто-

му связь между этими величинами слабая, ошибку нельзя считать случайной величиной. Из-за большой величины среднего квадрата ошибка логарифмическая модель не подходит для аппроксимации числа пожаров.

В конце рассмотрим логистическую кривую. Метод наименьших квадратов дал результат

$$Y_m = \frac{422,4461}{1 + 0,637634 * \exp(0,07957 * [X - 2001])} \quad (11)$$

Среднее значение квадрата ошибки составило 27,7 (Табл. 8). Это значение больше чем в линейной (5) и параболической (8) моделях.

Коэффициент линейной корреляции Пирсона между ошибкой и номером года равен 0,02. Поэтому связь между этими величинами очень мала, ошибку можно считать случайной величиной.

В итоге проведенного исследования из рассмотренных функциональных зависимостей минимум среднего значения квадрата ошибки дала логистическая функция.

Отметим, что коэффициенты в функциях (5-11) были найдены с помощью программы Microsoft Excel. При этом использовали средство Поиск решения. Приведенные значения коэффициентов давали минимум среднего значения квадрата ошибки между фактическими и модельными значениями.

Таблица 8. Логистическая модель

Год	Y	Y _m	ε	ε ²
2001	246,5	258,0	-11,5	131,4
2002	260,8	249,9	10,9	118,7
2003	239,2	241,7	-2,5	6,4
2004	233,2	233,5	-0,3	0,1
2005	229,8	225,1	4,7	21,9
2006	220,5	216,7	3,8	14,2
2007	212,6	208,3	4,3	18,2
2008	202,2	199,9	2,3	5,1
2009	187,6	191,6	-4,0	15,8
2010	179,5	183,3	-3,8	14,3
2011	168,5	175,1	-6,6	43,2
2012	162,9	167,0	-4,1	16,6
2013	153,5	159,0	-5,5	30,4
2014	150,8	151,2	-0,4	0,2
2015	145,9	143,6	2,3	5,4
2016	139,5	136,1	3,4	11,4
2017	133,1	128,9	4,2	17,7
среднее	192,1	192,3	-0,2	27,7

Отметим, что в практической работе трейдеров на фондовом и товарных рынках чаще всего исполь-

зуют линейные модели для выявления трендов [22]. Остальные математические функции для моделирования обстановки применяются редко в связи со сложностью математических вычислений.

Литература:

1. Пожары и пожарная безопасность в 2005 году: Статистический сборник. Под общей редакцией Н.П. Копылова. – М.: ВНИИПО, 2006. – 139 с.
2. Пожары и пожарная безопасность в 2006 году: Статистический сборник. Под общей редакцией Н.П. Копылова. – М.: ВНИИПО, 2007. – 137 с.
3. Пожары и пожарная безопасность в 2007 году: Статистический сборник. Под общей редакцией Н.П. Копылова. – М.: ВНИИПО, 2008. – 137 с.
4. Пожары и пожарная безопасность в 2008 году: Статистический сборник. Под общей редакцией Н.П. Копылова. – М.: ВНИИПО, 2009. – 137 с.
5. Пожары и пожарная безопасность в 2009 году: Статистический сборник. Под общей редакцией Н.П. Копылова. – М.: ВНИИПО, 2010. – 135 с.
6. Пожары и пожарная безопасность в 2010 году: Статистический сборник. Под общей редакцией В.И. Климкина. – М.: ВНИИПО, 2011. – 140 с.
7. Пожары и пожарная безопасность в 2011 году: Статистический сборник. Под общей редакцией В.И. Климкина. – М.: ВНИИПО, 2012. – 137 с.
8. Пожары и пожарная безопасность в 2012 году: Статистический сборник. Под общей редакцией В.И. Климкина. – М.: ВНИИПО, 2013. – 137 с.
9. Пожары и пожарная безопасность в 2013 году: Статистический сборник. Под общей редакцией В.И. Климкина. – М.: ВНИИПО, 2014. – 137 с.
10. Пожары и пожарная безопасность в 2014 году: Статистический сборник. Под общей редакцией А.В. Матюшина. – М.: ВНИИПО, 2015. – 124 с.
11. Пожары и пожарная безопасность в 2015 году: Статистический сборник. Под общей редакцией А.В. Матюшина. – М.: ВНИИПО, 2016. – 124 с.
12. Пожары и пожарная безопасность в 2016 году: Статистический сборник. Под общей редакцией Д.М. Гордиенко. – М.: ВНИИПО, 2017. – 124 с.
13. Государственный доклад о состоянии защиты населения и территории Российской Федерации от чрезвычайных ситуаций природного и техногенного характера в 2017 году [Электронный ресурс]. – М.: МЧС России. ФГБУ ВНИИ ГОЧС (ФЦ), 2018, 376 с. – Режим доступа: https://www.mchs.gov.ru/upload/site1/document_file/hniVNLexTC.pdf.
14. Shumway R., Stoffer D.S. Time Series Analysis and its Applications. – Springer, 2000. – 549 p.
15. Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M. Time Series Analysis: Forecasting and Control. - N.Y.: John Wiley and Sons, 2015. - 712 p.
16. say R.S. Analysis of financial time series. – N.Y.: Wiley, 2010. - 715 p.
17. Батура А.Н. Прогнозирование количества пожаров в регионе на основе теории временных рядов. Технологии гражданской безопасности, 2013, Т. 10, № 3 (37), с. 84-88.
18. Батура А.Н. Среднесрочное прогнозирование количества пожаров с использованием автокорреляционных функций. Природные и техногенные риски (физико-математические и прикладные аспекты), 2014, № 3 (11), с. 28-36.
19. Кайбичев И.А., Яковлев Е.Е. Корреляционный анализ основных показателей пожарной статистики в Российской Федерации за 2001-2015 годы // Актуальные вопросы естествознания: материалы II Межвузовской научно-практической конференции, Иваново, 12 апреля 2017 года / сост.: Н.Е. Егорова. – Иваново: Ивановская пожарно-спасательная академия ГПС МЧС России, 2017. – с. 158-161.
20. Кайбичев И.А., Калимуллина К.И. Регрессионный анализ основных показателей пожарной статистики в Российской Федерации // Актуальные проблемы обеспечения безопасности в Российской Федерации – с. 86-95. 22. Швагер Дж. Технический анализ. Полный курс. – М.: Альпина Паблишер, 2001. – 768 с.