

УДК 519.246.8

## ПРИМЕНЕНИЕ АДАПТИВНОГО РЕГРЕССИОННОГО МОДЕЛИРОВАНИЯ ДЛЯ ОПИСАНИЯ ФУНКЦИОНИРОВАНИЯ ТЕХНИЧЕСКОГО ОБЪЕКТА

© 2014 Т.Е. Родионова

Ульяновский государственный технический университет

Поступила в редакцию 29.06.2014

В работе рассмотрены математические модели описания состояния технического объекта, полученные методами регрессионного моделирования. Рассмотрены особенности показателей качества функционирования рассматриваемого объекта. Произведено сравнение моделей, полученных по выборкам разного объема. Используя внутренние меры качества, предложена регрессионная модель для оценки параметров технического объекта.

Ключевые слова: *регрессионное моделирование, технический объект, выборка, качество, модель, корреляция*

Требование безаварийного функционирования любого технического объекта является актуальной задачей. Для этого проводится регулярный контроль определенных показателей качества, причем значения входных и выходных параметров определяются по результатам измерений в режиме реального времени. Традиционно для оценки качества работы объекта используется набор показателей, имеющих определенный физический смысл (технические, физические, химические и т.д.). В работе рассматривается объект, который характеризуется набором  $p$  параметров  $x_1-x_p$ , среди которых можно выделить  $k$  управляемых факторов  $x_p, x_{p+1}-x_p$ . Измерялись также  $m$  показателей качества функционирования объекта  $y_1-y_m$ . Требовалось исследовать влияние регрессоров  $x_1-x_p$  на отклики  $y_1-y_m$ .

Для построения математической модели описанного объекта при  $p = 8, k = 2, m = 7$  был применен подход адаптивного регрессионного моделирования, подразумевающий исследование и выбор оптимальных методов. По совокупности имеющихся данных за год для каждого отклика  $y_1-y_7$  были построены и исследованы модели. Для расчетов использовался пакет Statistica [5]. Модели, полученные методом наименьших квадратов, характеризовались наличием незначимых параметров и эффектом мультиколлинеарности [1, 2]. Поэтому были исследованы возможности применения ряда методов адаптации к выявленным нарушениям основных предпосылок применения регрессионного анализа.

Для моделирования показателей качества объекта были получены регрессии с использованием выборок размером 30, 60, 100 дней и по ним произведены прогнозы на 2-4 следующих дня. Предлагаемые размеры выборки можно объяснить

тем, что при разделении исходного файла на выборки пришлось учитывать неизменность физико-химических показателей объекта в течение достаточно продолжительного времени. В частности, один из управляемых параметров изменяется в диапазоне от 0,12 до 0,22, и периоды неизменных значений могут быть от нескольких дней до сотни суток (за исследуемый период максимальное значение 113 дней). Известно, что наличие в обрабатываемом файле столбца из одинаковых значений при компьютерной обработке данных приводит к возникновению ошибки и остановке расчета. Ниже в качестве примера приведены модели множественной регрессии для откликов  $y_1, y_6$  полученные для выборок разного размера, а также значения коэффициента детерминации  $R^2$ .

Для годовой выборки:

$$y_1 = 22,93 + 0,02x_1 + 0,45x_2 - 0,02x_3 - 0,97x_4 + 1,11x_5 - 0,003x_6 - 3,73x_7 - 6,2x_8; R^2 = 0,16.$$

$$y_6 = 6,8 - 0,0003x_1 + 0,13x_2 - 0,008x_3 - 0,18x_4 - 0,06x_5 + 0,18x_6 - 1,24x_7 - 1,78x_8; R^2 = 0,35.$$

Для выборки из 100 наблюдений:

$$y_1 = 13,91 - 0,34x_1 + 0,77x_2 - 0,13x_3 + 0,21x_4 + 1,19x_5 + 0,26x_6 - 6,66x_7 + 31,95x_8; R^2 = 0,29$$

$$y_6 = -3,23 + 0,14x_1 - 0,07x_2 - 0,04x_3 - 0,41x_4 + 0,70x_5 + 0,11x_6 + 1,34x_7 + 2,48x_8; R^2 = 0,5.$$

Для выборки из 60 наблюдений:

$$y_1 = -8,68 + 1,17x_1 - 0,09x_2 - 0,19x_3 - 1,48x_4 + 4,35x_5 - 0,01x_6 + 2,71x_7 + 29,43x_8; R^2 = 0,44.$$

$$y_6 = -6,87 - 0,01x_1 - 0,3x_2 - 0,01x_3 - 1,25x_4 + 1,00x_5 + 0,15x_6 - 3,87x_7 + 1,72x_8; R^2 = 0,7.$$

Для выборки из 30 наблюдений:

$$y_1 = -282,8 + 2,32x_1 - 7,97x_2 - 0,09x_3 - 1,81x_4 + 0,77x_5 - 0,085x_6 + 83,25x_7 - 39,34x_8; R^2 = 0,31.$$

$$y_6 = 9,5 - 0,024x_1 - 0,27x_2 - 0,003x_3 + 0,05x_4 + 0,27x_5 + 0,003x_6 - 2,83x_7 + 2,46x_8; R^2 = 0,67.$$

Родионова Татьяна Евгеньевна, кандидат технических наук, доцент кафедры прикладной математики и информатики. E-mail: t.rodionova@ulstu.ru

Были получены модели множественной регрессии для всех откликов  $y_1$ - $y_7$ , которые являются показателями качества функционирования рассматриваемого объекта. Анализируя полученные модели для каждого отклика с учетом разного размера выборки можно говорить о неустойчивости полученных моделей в пространстве параметров. Об этом свидетельствует достаточно большой разброс оценок параметров по разным выборкам, в том числе и смена знака. Для оценки

качества полученных по выборкам моделей были использованы такие статистики (меры качества) как стандартная ошибка  $\sigma_{\Delta}$ , коэффициент детерминации  $R^2$ , общий  $F$ -критерий. Так как значения этих статистик получаются путем обработки данных, на основе которых строится модель, то они являются внутренними критериями. Данные меры для каждой из исследуемых моделей приведены в табл. 1.

**Таблица 1.** Внутренние меры качества моделей

Отклик	Объем выборки	$R^2$	F-критерий	Стандартная ошибка	Кол-во значимых регрессоров
$y_1$	30	0,31	1,16	0,73	1
	60	0,44	4,88	0,8	0
	100	0,29	4,54	0,89	0
	годовая	0,16	8,2	0,91	2
$y_2$	30	0,11	0,34	0,03	0
	60	0,47	5,57	0,02	2
	100	0,44	9,1	0,03	2
	годовая	0,28	16,22	0,04	3
$y_3$	30	0,25	0,8	0,05	0
	60	0,14	1,05	0,06	0
	100	0,18	2,5	0,06	1
	годовая	0,33	20,64	0,08	3
$y_4$	30	0,59	3,76	18,82	2
	60	0,49	5,74	15,99	3
	100	0,52	12,55	12,7	3
	годовая	0,12	6,19	20,99	5
$y_5$	30	0,15	0,47	0,06	0
	60	0,19	1,51	0,06	1
	100	0,15	2,03	0,06	1
	годовая	0,47	37,59	0,06	1
$y_6$	30	0,67	5,24	0,16	2
	60	0,7	15,25	0,2	2
	100	0,5	11,71	0,22	1
	годовая	0,35	23,09	0,3	1
$y_7$	30	0,35	1,41	0,02	1
	60	0,39	3,98	0,02	3
	100	0,31	5,07	0,02	3
	годовая	0,42	30,82	0,02	6

Анализируя полученные модели для всех откликов  $y_1$ - $y_7$  по внутренним мерам качества, можно сделать вывод, что размер выборки в 60 наблюдений является наиболее предпочтительным для построения прогноза.

Исследование корреляционной матрицы показало, что между регрессорами  $x_1$  и  $x_3$ ,  $x_1$  и  $x_8$ ,  $x_2$  и  $x_7$  существует сильная линейная взаимосвязь. Это явление называется мультиколлинеарностью и имеет весьма отрицательные последствия для оценивания коэффициентов регрессии [1]. Следует также отметить, что отклик  $y_4$  связан

корреляционной связью средней силы с  $x_7$ , а отклик  $y_6$  с регрессором  $x_6$ . В табл. 2 приведены коэффициенты парной корреляции между параметрами  $x_1$ - $x_8$  и откликами  $y_1$ - $y_7$  (набором показателей качества технического объекта).

Анализ остатков, полученных для моделей по 60 наблюдениям, показали, что для каждого из откликов существуют выбросы (значения, выходящие за стандартное отклонение остатков). В табл. 3 приведен список выбросов, обнаруженных при построении моделей для каждого из откликов.

**Таблица 2.** Коэффициенты парной корреляции

	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$	$y_7$
1	2	3	4	5	6	7	8
$x_1$	0,09	-0,13	-0,45	-0,15	-0,67	0,18	-0,48
$x_2$	0,25	0,40	0,09	0,14	0,45	0,13	0,31
$x_3$	-0,05	0,04	-0,16	0,11	-0,14	-0,03	-0,18
$x_4$	-0,04	0,11	0,06	-0,17	-0,16	0,12	-0,19

1	2	3	4	5	6	7	8
$x_5$	0,08	0,12	0,17	0,05	0,26	0,01	0,52
$x_6$	0,10	0,30	0,04	-0,18	-0,17	0,57	0,01
$x_7$	0,24	0,39	0,09	0,14	0,46	0,12	0,32
$x_8$	-0,09	0,15	0,40	0,14	0,66	-0,18	0,44

Таблица 3. Выявленные выбросы (аномальные наблюдения) для откликов

Отклик	Наблюдения с зафиксированными выбросами	Количество выбросов
$y_1$	42	1
$y_2$	14, 21, 26, 51	4
$y_3$	-	-
$y_4$	4, 27	2
$y_5$	9, 12, 36	3
$y_6$	19, 54	2
$y_7$	9, 28, 42	3

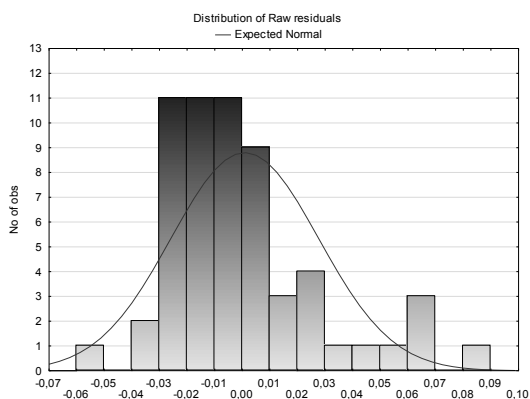


Рис. 1. Гистограмма остатков для отклика  $y_2$

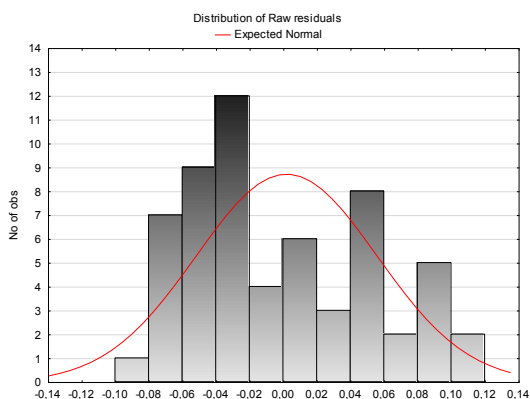


Рис. 2. Гистограмма остатков для отклика  $y_3$

На рис. 1 и 2 приведены гистограммы остатков для откликов  $y_2$  и  $y_3$ , которые позволяют сделать вывод о нарушении предположения регрессионного анализа о нормальности распределения остатков. Приведенные гистограммы остатков для исследуемых моделей получены с использованием пакета Statistica [5].

Для контроля рассматриваемого объекта с целью регулирования его параметров необходима модель, содержащая управляемые параметры  $x_7$  и  $x_8$ . Рассматривая в совокупности полученные меры качества моделей, и учитывая выявленные нарушения условий МНК, далее предложены некоторые методы адаптации [4, 6, 7]. Для преодоления

выявленных нарушений был использован метод пошаговой регрессии. Пошаговая множественная регрессия является статистическим методом анализа связи между зависимой переменной ( $y$ ) и множеством независимых переменных ( $x_1, x_2, \dots, x_p$ ) и осуществляет отбор независимых переменных в порядке их значимости. Критерий значимости основывается на уменьшении сумм квадратов. Независимая переменная, наиболее влияющая на это уменьшение на данном шаге, вводится в регрессию. Одним из существенных элементов данной реализации является добавление в алгоритм операции исключения регрессора, включенного в модель на данном шаге и ухудшающего значение критерия, по которому производится поиск оптимальной модели.

Далее для анализа был применен метод случайного поиска с адаптацией и случайного поиска с возвратом. Задачу поиска оптимального набора регрессоров можно рассматривать как задачу оптимизации функционалов с булевыми переменными  $z_j (j = \overline{1, p})$ , равными 0 или 1. Методы пошаговой регрессии и случайного поиска с адаптацией позволили получить модели со значимыми параметрами и хорошими прогностическими свойствами, но не все из них содержат необходимые управляемые параметры  $x_7$  и  $x_8$ . Значение коэффициента детерминации (0,67) в соответствии с принятой шкалой соответствует «заметной» силе связи. В состав моделей входит от 2 до 6 параметров объекта, при этом не во все модели вошли оба управляемых параметра  $x_7$  и  $x_8$ . Если не использовать принудительное введение этих параметров, то подходящей модели по методу пошаговой регрессии, как и по МНК, не получено. Гребневое оценивание позволило получить модели с малыми значениями коэффициента детерминации и F-критерия, что не соответствует требованиям оптимальной модели. Результаты расчетов методом пошаговой регрессии и гребневого оценивания приведены в табл. 4. В ней приводятся номера параметров, вошедшие в модель, полученную пошаговым методом и внутренние меры качества для каждого из примененных методов.

**Таблица 4.** Меры качества модели при пошаговом и гребневом оценивании

От кл ик	Пошаговая регрессия			Гребневая регрессия			
	№ параметра	$R^2$	$F$ -кри- терий	стандарт ная ошибка	$R^2$	$F$ -кри- терий	стандарт ная ошибка
$y_1$	1, 2, 3, 5	0,42	10,08	0,82	0,4	4,28	0,86
$y_2$	2, 5, 6, 7, 8	0,47	9,22	0,03	0,36	3,53	0,03
$y_3$	3, 5, 7	0,13	2,9	0,06	0,13	0,9	0,05
$y_4$	2, 3, 5, 6, 7, 8	0,47	9,56	15,6	0,27	2,33	18,9
$y_5$	1, 2, 7, 8	0,18	2,9	0,05	0,15	1,1	0,06
$y_6$	2, 3, 4, 5, 6, 7	0,7	20,95	0,2	0,7	12,69	0,21
$y_7$	2, 5, 7	0,28	7,0	0,2	0,3	2,6	0,02

При использовании метода случайного поиска с адаптацией, были получены модели, характеризующиеся небольшой стандартной ошибкой, сравнимой по значениям с результатами множественной и пошаговой регрессии. По коэффициенту детерминации и  $F$ -критерию получено небольшое улучшение качества модели. Основным достоинством данных моделей можно считать тот факт, что каждая из них содержит управляемые параметры  $x_7$  и  $x_8$ , которые необходимы для регулирования состояния технического объекта.

При исследовании влияния объема выборки на характеристики полученных моделей, используя только внутренние меры качества, был выбран наиболее предпочтительный объем, равный 60 измерениям. Анализ корреляционных связей между параметрами модели и показателями качества рассматриваемого технического объекта позволил выявить существенные взаимосвязи, которые требуют дальнейшего исследования. При исследовании остатков полученных моделей были зафиксированы аномальные наблюдения, которые требуется исключить из исходных данных при дальнейшей обработке. Методы пошаговой регрессии и случайного поиска с адаптацией дали модели с хорошими прогностическими способностями, но не полностью соответствующие заявленным требованиям.

Дальнейшие исследования предполагают получение моделей с максимально возможным набором параметров по выборкам объемом 60 наблюдений. Выбор наилучшей для прогноза модели предполагает использование различных мер качества моделей, как внутренних, так и внешних. Для адаптации к выявленным нарушениям планируется использовать также методы гребневого и робастного оценивания.

**СПИСОК ЛИТЕРАТУРЫ:**

1. *Валеев, С.Г.* Анализ методов оценки параметров при мультиколлинеарности переменных / *С.Г. Валеев, Т.Е. Родионова* // Известия Вузов. Серия: Геодезия и аэрофотосъемка. 1999. №5. С. 20-28.
2. *Валеев, С.Г.* Методика статистической обработки РСДБ-наблюдений / *С.Г. Валеев, Т.Е. Родионова, В.Е. Жаров* // Известия Вузов. Серия: Геодезия и аэрофотосъемка. 2008. №1. С. 13-18.
3. *Клячкин, В.Н.* Информационно-математическая система раннего предупреждения об аварийной ситуации / *В.Н. Клячкин, Ю.Е. Кувайскова, А.А. Аleshina, Ю.А. Кравцов* // Известия Самарского научного центра РАН. 2013. №4(4). С. 919-923.
4. *Рыбкина, М.В.* Анализ зависимости качества жизни от развития социальных структур / *М.В. Рыбкина, Т.Е. Родионова* // Сборники конференций НИЦ Социосфера. 2013. №51. С. 051-053.
5. *Халафян, А.А.* STATISTICA 6. Статистический анализ данных. 3-е изд. – М.: ООО «Бином-Пресс», 2007. 512 с.
6. *Родионова, Т.Е.* Исследование взаимозависимостей между социально-экономическими показателями методами регрессионного анализа данных / *Т.Е. Родионова, М.В. Рыбкина* // Региональная экономика: актуальные вопросы и новые тенденции: сборник науч. трудов Междун. заочной науч.-практ. конф. (23-25 апреля 2014 года, г. Ульяновск). В 2 т. – Ульяновск: УлГТУ, 2014. Т. 2. С. 196-202.
7. *Родионова, Т.Е.* Сравнение регрессионных моделей показателей качества питьевой воды // Матер. 3-й науч.-практ. интернет-конференции 20-21 февраля 2014, Тольятти «Междисциплинарные исследования в области математического моделирования и информатики» – Тольятти, 2014. С. 159-162.

**APPLICATION OF ADAPTIVE REGRESSION MODELING FOR DESCRIPTION THE TECHNICAL OBJECT FUNCTIONING**

© 2014 T.E. Rodionova  
Ulyanovsk State Technical University

In work mathematical models of the description the state of technical object which are come into by methods of regression modeling are considered. Features of indicators the quality of functioning of the considered object are considered. Comparison of the models received on selections the different volume is made. Using internal measures of quality, the regression model for estimation the parameters of technical object is offered.

Key words: regression modeling, technical object, selection, quality, model, correlation

*Tatiana Rodionova, Candidate of Technical Sciences, Associate Professor at the Department of Applied Mathematics and Computing Science. E-mail: t.rodionova@ulstu.ru*