Alternative Methods for Constructing of Linear Regressions

M.O. Onizhuk, V.V. Ivanov*, A.V. Panteleimonov, Yu.V. Kholin

V.N. Karazin Kharkiv National University, 4 Svobody Sq., Kharkiv, 61022, Ukraine

Recieved: May 30, 2017; Accepted: October 17, 2017
DOI: 10.17721/moca.2017.105-111

Different approaches to solution of linear regression problem have been implemented and tested with Python programming language. The described methods are based on two concepts: the minimization of deviation of calculated value from experimental one of the dependent variable (methods OLS, Ordinary Least Squares, and LAD, Least Absolute Deviation) and the minimization of the errors perpendicular to the regression line (methods ODR, Orthogonal Distance Regression, and LADOD, Least Absolute Deviation of Orthogonal Distances). The significant influence of the chosen function on the regression equation has been demonstrated. The set of examples, including the model problems as well as the experimental datasets, shows the importance of appropriate selection of regression model when the spread in data is significant. In such a situation, the proposed in article new method LADOD "Least absolute deviations of Orthogonal Distances" demonstrated adequate description of the system.

Keywords: QSAR, regression, least square method, least absolute deviation, orthogonal distance regression

Альтернативные методы построения линейной регрессии

Н.О. Онижук, В.В. Иванов*, А.В. Пантелеймонов, Ю.В. Холин

Харьковский национальный университет имени В. Н. Каразина, площадь Свободы 4, Харьков, Украина, 61022
*e-mail: vivanov@karazin.ua

Поступила: 30 мая 2017 г; Принята: 17 октября 2017 г
DOI: 10.17721/moca.2017.105-111

На алгоритмическом языке Python программно реализованы и тестированы различные подходы к решению задачи нахождения уравнений линейной регрессии. Рассмотренные методы построены как на минимизации отклонения зависимой переменной от вычисленного значения (методы OLS, Ordinary Least Squares, и LAD, Least Absolute Deviation), так и на минимизации длины перпендикуляра от экспериментальной точки до регрессионной прямой (методы ODR, Orthogonal Distance Regression, и LADOD, Least Absolute Deviation of Orthogonal Distances). Продемонстрировано существенное влияние вида минимизируемой функции на результирующее уравнение. Ряд примеров, включающий как чисто модельные задачи, так и экспериментальные наборы данных, показывает важность отбора регрессионной модели в ситуациях, которые характеризуются большим «разбросом» величин исходных данных. При этом вполне адекватным оказывается предложенный в работе метод "наименьших модулей ортогональных расстояний" LADOD.

Ключевые слова: QSAR, регрессия, метод наименьших квадратов, метод наименьших модулей, регрессия ортогональных расстояний

Регрессионный анализ является типичным инструментом в решении задач физической и аналитической химии, а также при построении статистических моделей типа «структура-свойство» (*Quantity-Structure Activity Relationship*, **QSAR**). Основным математическим аппаратом при построении регрессионных уравнений является стандартный метод наименьших квадратов (*Ordinary Least Squares*, **OLS**) [1,2]. Ряд модификаций **OLS** связан с преодолением осложняющих расчет обстоятельств. Среди таких обстоятельств мультиколлинеарность исходных данных, избыточность предикторного набора и неустойчивость результатов расчета относительно отдельных выбросов (не-робастность **OLS**). Среди обобщений **OLS** – взвешенный метод наименьших квадратов, методы регуляризации (**Ridge**-регрессия эквивалентна методу регуляризации по Тихонову [3], **LASSO** [4]), регрессия на главных компонентах (**PCR**), «неполный» метод наименьших квадратов (**PLS**) [5,6] и *т.д.*

Следует отметить, что геометрический смысл всех перечисленных подходов остается неизменным – минимизируются отклонения зависимой переменной (y) от ее вычисленного по регрессии значения $y^{(calc)}$, при заданных величинах предикторов (x). Разумеется, в тех ситуациях, когда исходные данные достаточно точно соответствуют линейной гипотезе, а минимизируемая задача приводит к хорошо обусловленной системе уравнений, минимизация $y - y^{(calc)}$, в духе **OLS,** дает однозначное решение задачи о линейной регрессии. Однако, в практических расчетах, часто встречаются ситуации, когда исходные данные осложнены наличием значительной случайной погрешности, что проявляется как в предикторном наборе (*x*) так и в наборе зависимых переменных (*y*). В этом случае результирующее уравнение (его регрессионные параметры) могут быть чувствительными к методу вычисления. Целью настоящей работы является проведение модельных расчетов, основанных на различных альтернативных способах построения линейной функции. При этом мы рассматриваем проблему с прагматической (эвристической) точки зрения – т.е. априорно предполагаем существование робастного решения задачи, не заботясь о статистическом обосновании модели. Такая постановка задачи вполне типична в современных построениях **QSAR.** Соответствующий программный комплекс реализован нами на скриптовом языке **Python**.

Методы нахождения уравнений регрессии

Для простейшего варианта регрессии:

$$y = \beta_0 + \beta_1 x \qquad (1)$$

графическая иллюстрация способа построения соответствующего уравнения дана на рис. 1а.

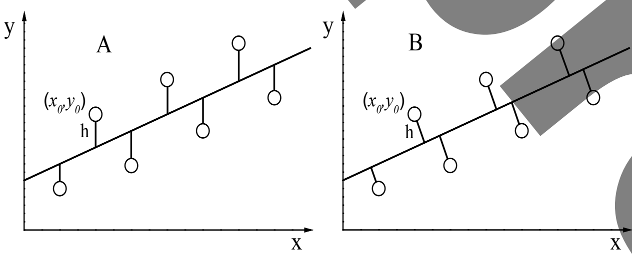

Рис. 1. Графическое представление стандартного (**OLS**) метода построения регрессионного уравнения (а) и метода **ODR** (в).

Здесь расстояние от точки с координатами (x_0, y_0) до прямой $y - \beta_0 + \beta_1 x = 0$, вдоль линии параллельной оси «у» равно (рис. 1а):

$$h = y_0 - \beta_0 + \beta_1 x_0. \qquad (2)$$

Исходя из (2), в методе **OLS** минимизируется величина h^2 . Определенной разновидностью метода, в рамках представления рис. 1а, является метод наименьших модулей (*Least Absolute Deviation*, **LAD)**, в котором уравнение строится путем минимизации соответствующего модуля, |h|.

Однако, с геометрической точки зрения, корректное расстояние от точки до прямой (1) определяется перпендикуляром (рис. 1в):

$$h = \pm \frac{|y_0 - \beta_0 - \beta_1 x_0|}{\sqrt{1 + \beta_1^2}} . \qquad (3)$$

В современной научной литературе нет единства в обозначении метода построения регрессионной линии, основанной на формуле (3). Мы будем обозначать его **ODR** (*Orthogonal Distance Regression*). В этом случае минимизируются суммы квадратов расстояний h^2 согласно (3). Для наиболее общих вариантов этого метода распространены следующие названия **TLS** (*Total Least Squares*), *Confluent Analysis*, **EIV** (*Error-in-Variables regression*) [7-9]. В последней аббревиатуре подчеркивается характерная особенность **ODR** – предположение о наличии погрешностей как в зависимых переменных, так и в предикторах.

Отметим, что возможна также реализация **ODR** и в рамках метода наименьших модулей (**LAD)**, что соответствует минимизации суммы модулей расстояний, |h| ,определенных согласно формуле (3). Такой вариант расчета ранее не был исследован, мы будем называть его **LADOD** (метод наименьших модулей ортогональных расстояний, *Least Absolute Deviation of Orthogonal Distances*).

В целом, для регрессии (1), минимизируемые функции различных, выше перечисленных, методов имеют вид:

$$U_{LS}(\beta_0, \beta_1) = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2 , \qquad (4)$$

$$U_{LAD}(\beta_0, \beta_1) = \sum_i |y_i - \beta_0 - \beta_1 x_i| = \sum_i \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{w_i} , \qquad (5)$$

$$w_i = |y_i - \beta_0 - \beta_1 x_i| , \qquad (6)$$

$$U_{ODR}(\beta_0, \beta_1) = \frac{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}{1 + \beta_1^2} , \qquad (7)$$

$$U_{LADOD}(\beta_0, \beta_1) = \frac{\sum_i |y_i - \beta_0 - \beta_1 x_i|}{\sqrt{1 + \beta_1^2}} . \qquad (8)$$

Приравнивая соответствующие производные к нулю:

$$\partial U(\beta_0, \beta_1) / \partial \beta_0 = 0 , \quad \partial U(\beta_0, \beta_1) / \partial \beta_1 = 0 \qquad (9)$$

можно получить выражения для коэффициентов регрессии. При этом, коэффициент β_1, в различных моделях, удобно выразить через соответствующие *к*овариации s_{xy} и средние значения (Табл. 1). Для указанных моделей регрессии, коэффициент β_0 вычисляется одинаково:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}. \qquad (10)$$

Следует отметить, что наличие варьируемых весов (w_i) в методе **LAD**, не позволяет получить замкнутые выражения для регрессионных коэффициентов, которые зависели бы лишь от x_i и y_i. Для реализации такого подхода ранее было предложено ряд алгоритмов [10,11]. В настоящей работе мы будем следовать методу «*вариационно-взвешенных квадратичных приближений*» [12]. Согласно этому подходу решения **LAD** (и **LADOD)** можно получить, используя итерационный процесс по схеме:

$$\underset{\text{(начальное приближение)}}{\beta_0, \beta_1} \to w_i \to \tilde{x}, \tilde{y} \to \tilde{s}_{xx}, \tilde{s}_{xy} \to \beta_0, \beta_1\,_{new} \quad (11)$$

Процесс прерывается при условии малости изменений искомых параметров регрессии на соседних итерациях:

$$\left|\beta_0^{(k+1)} - \beta_0^{(k)}\right| + \left|\beta_1^{(k+1)} - \beta_1^{(k)}\right| = \eta_k \le \varepsilon\,, \quad (12)$$

где η – достаточно малая величина, а k – номер итерации. При этом, чтобы избежать сингулярности в выражении (5) в случае $w_i \approx 0$, мы использовали весовые коэффициенты в виде:

$$w_i = |y_i - \beta_0 - \beta_1 x_i| + 0.01\eta_k. \quad (13)$$

Таким образом, для нахождения уравнения регрессии методом **ODR** необходимо решение квадратного уравнения, в то время как для метода **LADOD,** на каждой стадии итерационного процесса (11), необходимо решать кубическое уравнение.

Таблица 1. Выражения для регрессионных коэффициентов и вспомогательных величин в различных подходах.

Метод	Средние значения	Ковариация s_{xy}	Коэффициент β_1
OLS	$\bar{x} = \frac{\sum x_i}{n}$	$s_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$	$\beta_1 = \frac{s_{xy}}{s_{xx}}$
ODR			$s_{xy}\beta_1^2 + (s_{xx} - s_{yy})\beta_1 - s_{xy} = 0$
Коэффициент $\beta_0 = \bar{y} - \beta_1\bar{x}$;			
LAD	$\tilde{x} = \frac{\sum x_i \cdot w_i^{-1}}{\sum w_i^{-1}}$	$\tilde{s}_{xy} = \sum \frac{(x_i - \tilde{x})(y_i - \tilde{y})}{w_i}$	$\beta_1 = \frac{\tilde{s}_{xy}}{\tilde{s}_{xx}}$
LADOD			$\tilde{s}_{xx}\beta_1^3 + (2\tilde{s}_{xx} - \tilde{s}_{yy})\beta_1 - 2\tilde{s}_{xy} = 0$
Коэффициент $\beta_0 = \tilde{y} - \beta_1\tilde{x}$			

Тестовые расчеты

Для изучения свойств регрессионных моделей мы рассмотрели серию тестовых выборок. Для этого брался набор из 18 точек (x_i, y_i) соответственно уравнению:

$$y_i = 2x_i - 3. \quad (14)$$

Затем, в значения x_i и y_i, вносились нормально распределенные погрешности с различной дисперсией (Табл. 2). В работе проводилось сравнение адекватности регрессионных уравнений в двух случаях. В первом случае значения предиктора точно известны, а значения зависимой переменной (y) содержат погрешность (стандартное допущение при использовании метода наименьших квадратов). Во втором случае как x, так и y имеют сопоставимую погрешность. Соответствующий коэффициент корреляции переменных x и y (по Пирсону, r) характеризует отклонение точек данных от линии (14).

Таблица 2. Коэффициенты уравнений регрессии в различных подходах.

Отклонения	σ	r	**OLS** β_0	**OLS** β_1	**LAD** β_0	**LAD** β_1	**ODR** β_0	**ODR** β_1	**LADOD** β_0	**LADOD** β_1
	0.1	0.998	-3.07	2.00	-3.04	2.00	-3.07	2.00	-3.04	2.00
	0.5	0.999	-2.77	1.99	-2.47	1.97	-2.80	1.99	-2.90	2.00
y	1	0.996	-1.97	1.90	-1.31	1.85	-2.12	1.91	-2.11	1.94
	3	0.960	-2.27	2.06	-1.40	1.92	-4.00	2.22	-3.88	2.29
	3	0.959	-3.96	2.17	-2.98	2.02	-4.10	2.20	-2.58	1.90
	0.1	0.999	-2.96	1.99	-2.91	1.99	-2.97	1.99	-2.91	1.987
	0.5	0.996	-3.28	2.00	-3.07	2.01	-3.40	2.01	-3.07	2.01
x, y	1	0.981	-4.31	2.11	-2.68	1.96	-5.02	2.18	**-6.30**	2.24
	3	0.862	-1.39	1.93	**2.96**	1.53	**-7.08**	2.46	-3.85	2.04
	3	0.808	**2.83**	1.48	**3.64**	1.58	-3.35	2.07	**-10.3**	2.71
Теоретические значения β_0 = -3.000, β_1 = 2.000										

Как видно из приведенных значений, при достаточно высокой точности описания y (s = 0.1, 0.5) все регрессионные модели показывают вполне адекватный результат. Наиболее близкими к идеальной линии оказались результаты метода **LADOD.** Напомним, что каждая из перечисленных регрессионных моделей оптимальна в своем смысле (уравнения 4-9). При значительных величинах стандартного отклонения s = 1, 3 (для s = 3 формировались две различных выборки) наиболее близкий результат к теоретически заданному показывает либо **LADOD** (Рис. 2*а*), либо **ODR** (Рис. 2в).

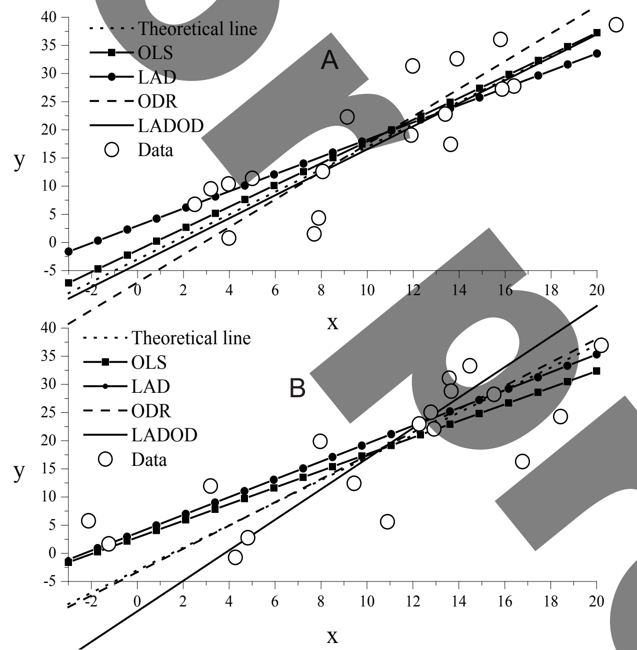

Рис. 2. Аппроксимация тестовых выборок, содержащих ошибки (σ = 3) как в *x*, так и в *y.*

Обратим, однако, внимание на то, что в некоторых случаях величина β_0 обладает большой погрешностью и даже имеет неверный знак (в Табл. 2 выделено черным), что, по-видимому, связано с потенциальной множественностью решений методов **LAD** и **LADOD**. При этом, даже в случае, когда погрешность вносится только в значения зависимой переменной достаточно сложно однозначно указать лучший подход. При большой погрешности (s = 3) более близкой к идеальной линии (14) оказывается результат **LAD** (рис. 3). В данном случае подходы, построенные на минимизации перпендикуляра к прямой, по очевидным причинам, дают более грубый результат.

Таким образом, методы, основанные на идеологии **LAD,** могут оказаться более точными, но при малом кол-ве точек (линия **LAD** всегда проходит через две точки) результат может оказаться бессмысленным (слишком грубым). Метод **ODR**, в целом, при малых величинах s, качественно верно формирует функцию, хоть и приводит иногда к существенным отклонениям от идеальных значений при больших значения s.

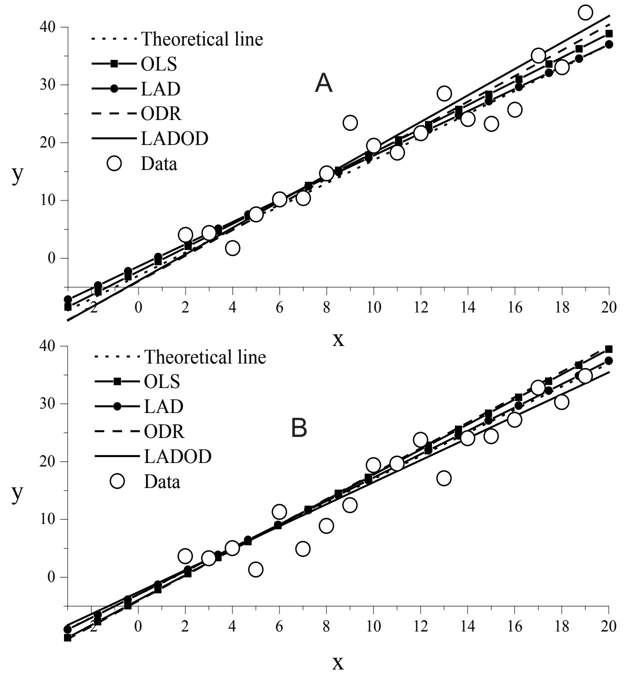

Рис. 3. Аппроксимация тестовых выборок, содержащих ошибки (σ = 3) только в *y*.

Валидация полученных результатов, для выборок с наибольшей дисперсией, была проведена с помощью процедуры *leave-one-out* (**LOO**). Критерии, которые характеризуют корректность и предсказательную способность моделей (коэффициенты детерминации), рассчитывались по известным в хемометрии формулам (см. например [13]):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_{i\,calc})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \quad Q^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_{i\,pred})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \qquad (15)$$

где y_i – исходные (аппроксимируемые) величины, $y_{i\,calc}$ – рассчитанные, а $y_{i\,pred}$ – величины предсказанные по процедуре **LOO**. Заметим, что хорошей считается модель с достаточно малой величиной η = R^2 - Q^2 [14]. Для коэффициента корреляции по Пирсону мы сохраняем символ ***r*.** Полученные величины критериев указаны в Табл. 3.

Падения величины Q^2 по сравнению с R^2, дает возможность сравнивать робастность и прогностическую способность моделей. Расчеты показывают неплохое качество прогноза для ситуации, когда погрешность вносится лишь в y. Максимальная разность η = R^2 - Q^2 достигается в методах **OLS** (η = 0.025) и **LADOD** (η = 0.05), что весьма неплохо. Но удивительно, что когда погрешность вносится и в x, и в y, метод **ODR** демонстрирует наихудшую прогностическую способность (Q^2 = 0.405; η = 0.143)! Во всех случаях наиболее стабильные результаты показывает метод наименьших модулей (**LAD**).

Таблица 3. Значения статистических критериев, характеризующих регрессионные модели для двух выборок с σ = 3.

Отклонения	r	OLS		LAD		ODR		LADOD	
		R^2	Q^2	R^2	Q^2	R^2	Q^2	R^2	Q^2
y	0.960	0.911	0.886	0.904	0.901	0.905	0.880	0.904	0.854
	0.959	0.925	0.907	0.921	0.913	0.921	0.901	0.922	0.913
x,y	0.862	0.743	0.692	0.710	0.688	0.682	0.595	0.727	0.583
	0.808	0.653	0.568	0.621	0.577	0.548	0.405	0.621	0.523

Типичной модельной задачей является ситуация, когда одна или несколько точек значительно отклоняются от линии регрессии (выпадают), Рис. 4.

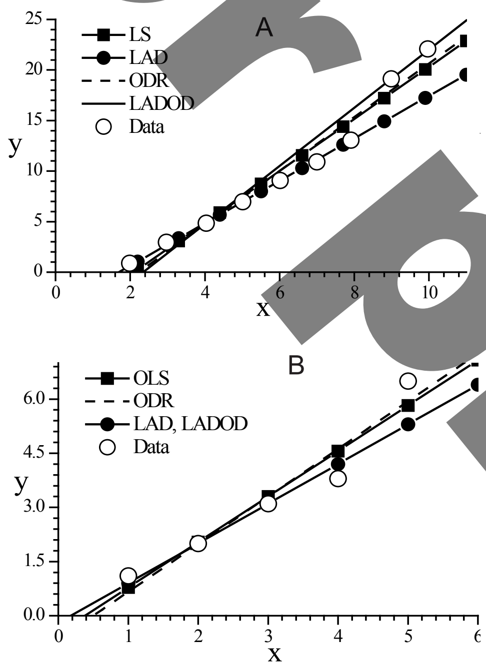

Рис. 4. Модельная задача с двумя (A) и с одной (B) «выпадающими» точками.

В этом случае хорошо видна не-робастность метода **OLS** по сравнению с **LAD**. Интересно, что две выпадающие точки оказали наибольшее влияние на уравнение, полученное методом **LADOD**. Однако процедура **LOO** показала, что выброс одной точки может привести к сильному изменению наклона аппроксимирующей прямой в методах, построенных на минимизации суммы модулей. В задаче с малым количеством точек это приводит к нестандартной ситуации, когда величина η в методах **OLS** и **ODR** может оказаться даже меньше, чем в методе **LAD,** Табл. 4. В случае обработки классических модельных данных, описанных в работах [15,16], метод **LADOD** показал достаточно высокую предсказательную способность. Величина $\eta = 0.010$ для **LADOD** заметно больше, чем для метода **LAD** (Табл. 5).

Таблица 4. Коэффициенты регрессии и статистические критерии для набора с двумя выпадающими точками.

Метод	Коэффициенты регрессии		R^2	Q^2	r
OLS	β_0 β_1	-5.41 2.57	0.963	0.932	0.031
LAD	β_0 β_1	-3.56 2.10	0.910	0.842	0.068
ODR	β_0 β_1	-5.92 2.66	0.962	0.931	0.031
LADOD	β_0 β_1	-6.75 2.88	0.945	0.907	0.038

Таблица 5. Коэффициенты регрессии и статистические критерии для набора с одной выпадающей точкой из работ [15,16].

Метод	Коэффициенты регрессии		R^2	Q^2	r
OLS	β_0 β_1	-0.480 1.260	0.931	0.720	0.211
LAD	β_0 β_1	-0.200 1.100	0.904	0.684	0.220
ODR	β_0 β_1	-6.56 1.319	0.929	0.695	0.234
LADOD	β_0 β_1	-0.200 1.100	0.904	0.894	0.010

Еще одна важная модельная ситуация соответствует слабо коррелированному набору данных (Рис. 5). Глядя на эти данные вполне можно выдвинуть предположение о том, как должна была бы пройти такая линия, но результаты расчетов показали, что четыре рассмотренных метода дают два различных типа решений. Одно из них соответствует более пологому углу наклона аппроксимационной прямой. Это методы **OLS** и **LAD**. Второй тип решений соответствует бóльшему развороту линии к оси y (методы **ODR** и **LADOD**). Соответствующие параметры приведены в Табл. 6. Таким образом, нахождение однозначной линейной модели для такой задачи не представляется возможным.

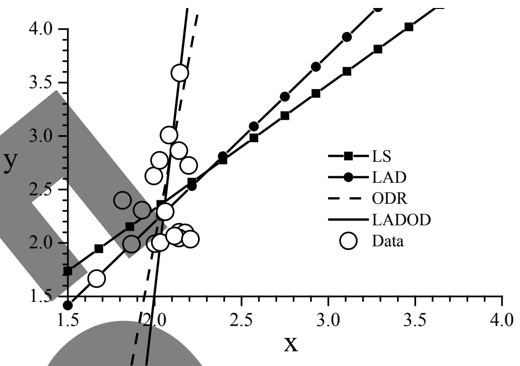

Рис. 5. Аппроксимация слабо коррелированных величин.

И, наконец, в последнем примере рассматривается задача о линейной связи температур кипения (T_{bp} ˚С) и плавления (T_{mp} ˚С) органических веществ. Соответствующие данные (всего 243 соединения) были взяты нами из [17]. Поскольку оба набора T_{bp} и T_{mp} очевидно имеют некую экспериментальную погрешность измерения, то можно было бы предположить, что **ODR** (и **LADOD**) должны быть наиболее точными. Однако результаты расчетов (см. табл. 7 и рис. 6), приводят к противоположным выводам. Методы **ODR** и **LADOD** дают большую величину углового коэффициента, β_1. Согласно процедуре **LOO**, методы **OLS** и **LAD** демонстрируют заметно лучшие и согласующиеся друг с другом результаты. Замечательно при этом, что величины η малы во всех методах.

Таблица 6. Коэффициенты регрессии для набора слабо коррелированных чисел, r = 0.350.

Коэффициенты регрессии	Метод			
	OLS	**LAD**	**ODR**	**LADOD**
β_0	-0.01	-0.93	-15.3	-26.7
β_1	1.16	1.56	8.7	14.1

Таблица 7. Коэффициенты регрессии зависимости температур кипения от температур плавления, r = 0.794.

Метод	Коэффициенты регрессии		R^2	Q^2	r
OLS	β_0	169.7	0.631	0.624	0.007
	β_1	0.96			
LAD	β_0	171.0	0.630	0.629	0.001
	β_1	0.97			
ODR	β_0	177.0	0.565	0.556	0.0
	β_1	1.27			
LADOD	β_0	179.8	0.580	0.580	0.0
	β_1	1.228			

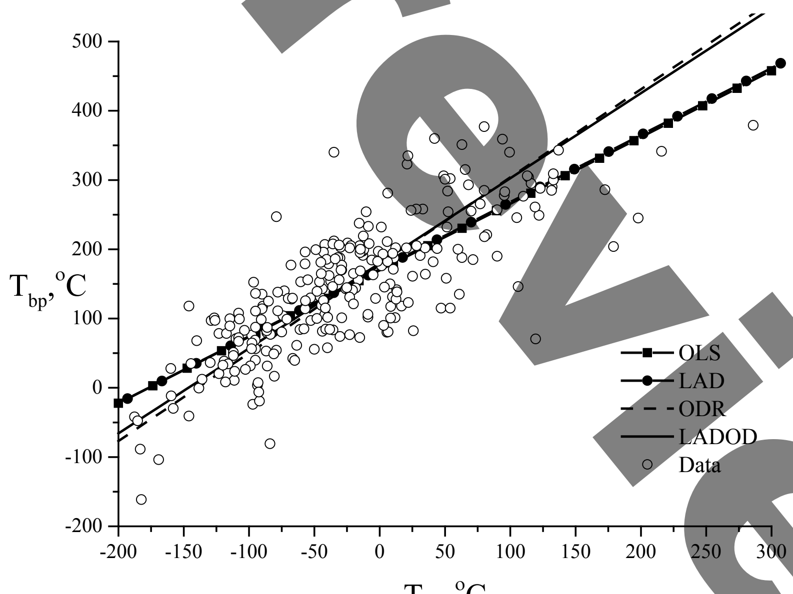

Рис. 6. Температуры кипения и плавления органических соединений.

Заключение

Различные методы построения уравнений регрессии были сформулированы практически одновременно еще на заре статистики как науки (см., например, [18]). Однако значительные вычислительные (и алгоритмические) сложности привели к тому, что в течении последующих двухсот лет почти все расчеты регрессий проходили методом наименьших квадратов. Лишь в последние десятилетия, в связи с интенсивным развитием компьютерной техники, внимание практиков привлекли и альтернативные подходы. Тем не менее, на сегодняшний день, в распространенных статистических пакетах методы **LAD**, **ODR** все еще остались не реализованными и, таким образом, возможности указанных подходов в хемометрических расчетах мало исследованы. В настоящей работе мы частично восполнили этот пробел. На языке **Python** нами разработан комплекс соответствующего программного обеспечения. Расчеты модельных систем показали, что различные подходы приводят в целом к эквивалентным результатам в случае, если разброс в исходных данных не слишком велик.

Метод **LAD** подтвердил свою робастность в задачах, когда число точек достаточно велико. Однако в противоположенном случае, если число точек мало, а разброс данных велик, **LAD**-решения могут оказаться неадекватными.

Методы **ODR** и **LADOD** в целом ведут себя похожим образом, и в ряде случаев показали хорошие результаты при больших разбросах в исходных данных. Однако их широкое применение все еще остается под вопросом.

В заключение отметим, что современный уровень вычислительной техники создал возможность использования одновременно нескольких методов построения регрессионных уравнений. Появление существенных различий в решениях может служить индикатором для выявления осложняющих обстоятельств и необходимости тщательного статистического анализа задачи.

Литература

1. Демиденко Е.З. Линейная и нелинейная регрессии. Москва: Финансы и статистика, 1981. С. 304.
2. Лоусон Ч., Хенсон Р. Численное решение задач метода наименьших квадратов. Москва: Наука, 1986. С. 232.
3. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. Москва: Наука, 1986. С. 288.
4. Tibshirani R. *J. Roy. Statist. Soc.* 1996. B 58, № 1. P. 267-288.
5. Wold S., Sjöström M. Eriksson L. *Chem. and Intell. Lab. Systems.* 2001. 58. P. 109.
6. Vinzi E., Chin W.W., Henseler J., Wang H. (Editors). Handbook of Partial Least Squares. Springer-Verlag: New York, 2010. P. 850.
7. Ahn S.J. Least Squares Orthogonal Distance Fitting of Curves and Surfaces in Space. Springer-Verlag: Berlin, Heidelberg, 2004. P. 127.
8. Van Huffel S., Vandewalle J. The total least squares problem. Computational Aspects and Analysis. Society for Industrial and Applied Mathematics: Philadelphia, 1991. P. 300.
9. Van Huffel S., Lemmerling P. (Editors). Total Least Squares and Errors-in-Variables Modeling Analysis, Algorithms and Applications. Springer Science + Business Media: Dordrecht, 2003. P. 398.
10. Bloomfield P., Steiger W.L. Least Absolute Deviations. Theory, Applications and Algorithms. Boston: Birkhäuser, 1983. P. 351.
11. Wesolowski G.O. *Comm. Stat. – Simul. Comp.* 1981. B 10, № 5. P. 479.
12. Мудров В.И., Кушко В.Л. Метод наименьших модулей. Москва: Знание, 1971. С. 64.
13. Golbraikh A., Tropsha A. *J. Mol. Graph. and Mod.* 2002. 20. P. 269.
14. Veerasamy R., Rajak H., Jain A., Sivadasan S. *Int. J. Drug Design and Discovery.* 2011. 2, № 3. P. 511.
15. Назаренко А.Ю., Сухан В.В., Назаренко Н.А. *Зав. лаб.* 1991. 57, № 10. С. 63.
16. Холин Ю.В., Никитина Н.А., Пантелеймонов А.В., и др. Метрологические характеристики методик обнаружения с бинарным откликом. Харьков. 2008. С. 87.
17. Tables of Physical & Chemical Constants (16th edition 1995). 3.3 Properties of organic compounds. Kaye & Laby Online. Version 1.0 (2005) www.kayelaby.npl.co.uk.
18. Heyde C.C., Seneta E. (Editors) Statisticians of the centuries, New York: Springer-Verlag, 2001. P. 500.