

## МЕТОД НАИМЕНЬШИХ КВАДРАТОВ В ЗАДАЧЕ, НЕ ИМЕЮЩЕЙ РЕШЕНИЯ

А.М. Жукавин, студент

Российский экономический университет им. Г.В. Плеханова

(Россия, г. Москва)

DOI:10.24412/2500-1000-2022-7-2-167-171

**Аннотация.** В данной статье автором рассматривается потенциал использования метода наименьших квадратов с целью прогнозирования значений случайных величин, находящихся в линейной зависимости. Также автором раскрывается несостоятельность метода наименьших модулей для решения задач подобного рода и обосновывается идея о том, что при достаточном количестве наблюдений двух коррелируемых переменных можно максимально достоверно выявить закон причинно-следственной связи между ними. Для этого, в частности, приводится пример решения задачи о проведении прямой через облако точек, не лежащих на одной прямой.

**Ключевые слова:** прогноз, причинно-следственный закон, регрессор, результирующая величина, линейная зависимость, отклонение.

*Бессмыслица – искать решение, если оно и так есть.*

*Речь идет о том, как поступать с задачей,*

*которая решения не имеет.*

**Стругацкий А. Н.**

Во всякой науке ощутима потребность прогнозировать те или иные события, но это затруднительно, ведь мы живём в недетерминированном мире, где многое зависит от стохастических, то есть случайных явлений, таких как вспышки на солнце, демографические скачки, или изменения котировок на нефтегазовом рынке. Будь то прогноз курса валют или прогноз погоды – вопрос точности измерений и надёжности полученных данных по-прежнему остается одной из ключевых проблем математики.

Многие процессы – социальные и природные – подчиняются закону причинно-следственной связи. Зная характер этой связи, можно построить её математическую модель, и – помещая в неё независимые количественные переменные (такие как время или Лондонская межбанковская ставка предложения) – получить прогноз для интересующей величины: от расстояния смещения литосферных плит до прироста ВВП.

Но невозможно построить прогноз, не располагая данными, иллюстрирующими

характер существования исследуемой системы. Самый лучший способ узнать будущее – заглянуть в прошлое. Ссылаясь на закон больших чисел, можно утверждать, что при стремлении количества наблюдений к бесконечности, имея в распоряжении данные о динамике изменения некоторой результирующей величины в зависимости от регрессора (независимой переменной, влияющей на результирующую величину), можно предположить, что эта динамика сохранится в исходном виде. А значит, выбрав интересующий нас регрессор, удастся определить, как поведёт себя результирующая величина при условии, что есть основания предполагать наличие корреляции между этими двумя величинами.

Допустим, в серии экспериментов были получены данные о зависимости двух количественных показателей, где  $Y$  – результирующая величина, поведение которой необходимо спрогнозировать, а  $X$  – независимая переменная (регрессор), определяющая  $Y$ . Пример динамики представлен в таблице 1.

Таблица 1. Динамика регрессора X и результирующей переменной Y

i	1	2	3	4	5	6	7	8
$y_i$	2	3	4	5	6	7	9	12
$x_i$	7	10	6	17	14	12	18	20

Для наглядности представим имеющиеся данные графически: изобразим корреляционное поле, на котором виден двумер-

ный массив данных, где каждому значению X соответствует значение Y (рис. 1).

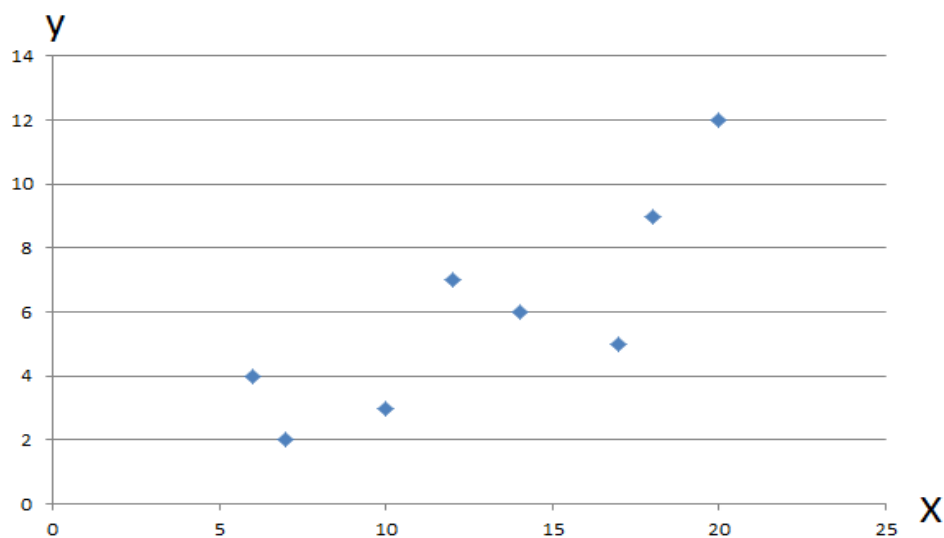


Рис. 1. Корреляционное поле регрессора X и результирующей переменной Y

Явно прослеживается некоторая корреляция и, если судить по виду облака точек, есть основания полагать, что данные находятся в прямой линейной зависимости. Это значит, что увеличение независимой переменной X влечёт за собой и увеличение результирующего показателя Y.

Так как мы наблюдаем рассеяние точек, то нельзя абсолютно точно сказать, как именно поведёт себя исследуемая система в дальнейшем. Но поскольку мы исходим из предпосылки о том, что характер взаимосвязи останется неизменным, для выявления причинно-следственного закона необходимо через эти точки провести прямую, которая послужит нам иллюстрацией динамики, и если выразить её в виде аппроксимирующей функции (аппроксимация – приближение), то она покажет, как именно меняется результирующий показатель в зависимости от регрессора. Такая задача является нерешаемой вследствие первой аксиомы принадлежности: провести такую прямую невозможно.

И, тем не менее, задачу необходимо решить. Следовательно, интересует ответ на

другой вопрос: какая прямая наиболее плотно примыкает к этой серии точек? Такую прямую найти можно, а наиболее плотное примыкание означает, что расстояние от значения на искомой прямой до значения на точке должно быть наименьшим. Данное расстояние называется отклонением наблюдаемых значений, а в связи с тем, что имеется не одна точка, а серия точек, задача сводится к следующему: сумма этих отклонений должна быть минимальной.

Поскольку мы предполагаем прямую линейную зависимость, а искомым графиком является прямая, становится очевидным то, что функция также должна быть линейного вида:

$$y = \alpha x + \beta.$$

Коэффициенты  $\alpha$  и  $\beta$  являются именно теми условиями, которые диктуют поведение результирующей переменной Y при изменении регрессора X. В тот момент, когда станет известно, каковы значения коэффициентов  $\alpha$  и  $\beta$ , задача будет решена. Если бы отклонения отсутствовали вовсе,

то значение каждого наблюдаемого случая составило бы:

$$\alpha x_i + \beta.$$

Но у нас также есть наблюдаемое значение  $y_i$ . Значит, отклонение — это разница между предполагаемым значением и наблюдаемым значением в каждой конкретной точке. Соответственно отклонение в точке  $X$  составит:

$$\alpha x_i + \beta - y_i.$$

Именно эти суммарные отклонения для каждой точки  $X$  мы хотим минимизировать.

Для достижения поставленной цели можно просто суммировать все отклонения, после чего подобрать коэффициенты  $\alpha$  и  $\beta$  таким образом, чтобы получившаяся сумма была минимальной. Проблема таится в том, что отклонение идёт в двух направлениях, вследствие чего может быть как положительным, так и отрицательным.

Во избежание взаимного уничтожения полученных значений и их обнуления, можно использовать метод наименьших модулей, так как по определению модуль — это расстояние от начала отсчёта до точки координатной прямой, соответствующей этому числу. Это поможет нивелировать

разницу в знаках, но сложность такой вычислительной процедуры высока, а выходные данные могут быть неверно интерпретированы. Также недостатком становится равнозначность для метода больших и маленьких отклонений при условии равенства их общей суммы. Поэтому надёжнее ликвидировать возможные отрицательные значения путём возведения отклонений в квадрат, чтобы наблюдаемое значение  $y_i$  сыграло свою отрицательную роль в каждой точке. Иначе говоря, рассмотрим сумму квадратов отклонений с первого наблюдения  $i=1$  до последнего наблюдения  $n$  (в нашем случае  $n=8$ ) включительно:

$$\sum_{i=1}^n (\alpha x_i + \beta - y_i)^2$$

Теперь вычислим квадраты отклонений для каждого наблюдаемого случая. Для этого в функцию:

$$\alpha x_i + \beta - y_i$$

подставим соответствующие значения  $y_i$  и  $x_i$  из таблицы 1, а затем возведём каждый отдельный результат в квадрат и таким образом, получим:

Таблица 2. Квадраты отклонений наблюдаемых значений

$i$	1	2	3	4
$(\alpha x_i + \beta - y_i)^2$	$(7\alpha + \beta - 2)^2$	$(10\alpha + \beta - 3)^2$	$(6\alpha + \beta - 4)^2$	$(17\alpha + \beta - 5)^2$
$i$	5	6	7	8
$(\alpha x_i + \beta - y_i)^2$	$(14\alpha + \beta - 6)^2$	$(12\alpha + \beta - 7)^2$	$(18\alpha + \beta - 9)^2$	$(20\alpha + \beta - 12)^2$

Ссуммируем полученные результаты:

$$f(\alpha; \beta) = (7\alpha + \beta - 2)^2 + (10\alpha + \beta - 3)^2 + (6\alpha + \beta - 4)^2 + (17\alpha + \beta - 5)^2 + (14\alpha + \beta - 6)^2 + (12\alpha + \beta - 7)^2 + (18\alpha + \beta - 9)^2 + (20\alpha + \beta - 12)^2.$$

Возникла функция, которая зависит от двух показателей  $\alpha$  и  $\beta$ . Когда же её значение станет минимальным, мы определим именно те коэффициенты  $\alpha$  и  $\beta$ , которые требуются для описания искомой прямой, следовательно, нужно оценить скорость изменения данной функции и исследовать

её на экстремумы. Для этого вычислим частные производные и найдем стационарные точки, то есть точки, в которых экстремум возможен. Продифференцировав функцию по переменной  $\alpha$ , а затем по переменной  $\beta$  и приведя подобные слагаемые, получим:

$$f'_\alpha = 3076\alpha + 208\beta - 1446;$$

$$f'_\beta = 208\alpha + 16\beta - 96.$$

Имеются две частые производные. Теперь требуется, чтобы обе обратились в ноль, ведь именно в этом случае первооб-

разная функция достигнет экстремума. В связи с этой необходимостью возникает система уравнений:

$$\begin{cases} 3076\alpha + 208\beta - 1446 = 0; \\ 208\alpha + 16\beta - 96 = 0. \end{cases}$$

Решением данной системы является:

$$\alpha = 0,53226;$$

$$\beta = -0,91935.$$

Выясняется, что сумма квадратов отклонений будет наименьшей именно при таких коэффициентах уравнения прямой, а значит, наиболее плотно к имеющимся

точкам будет прилегать прямая, являющаяся графиком функции прямой линейной зависимости:

$$y = 0,53226x - 0,91935.$$

Представим эту прямую на корреляционном поле.

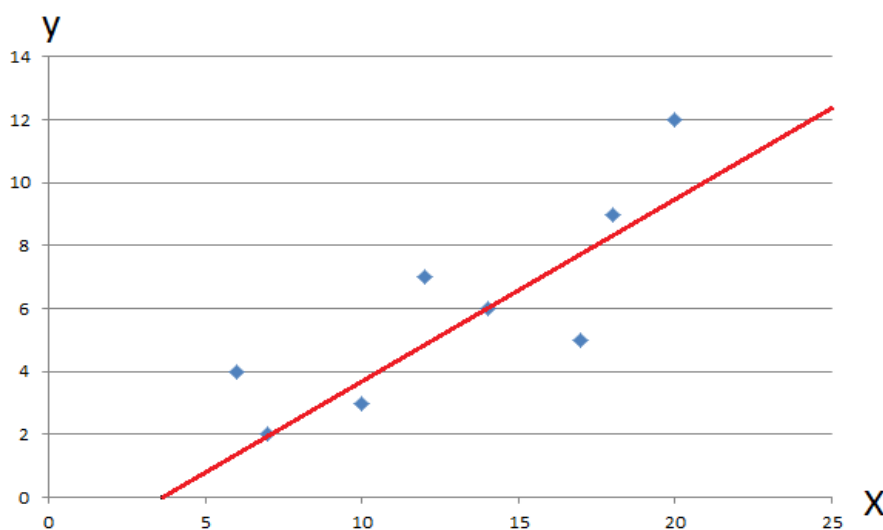


Рис. 2. График функции прямой линейной зависимости

Перед нами наиболее точная иллюстрация зависимости двух величин, которую можно выразить одной прямой: минимизировав суммарное отклонение, удалось максимизировать вероятность того, что в результате нового испытания случайная точка окажется именно на этой прямой или в её окрестностях.

Теперь вернёмся к изначальному вопросу: как спрогнозировать значение переменной  $Y$  в дальнейшем при конкретном значении переменной  $X$ ? Если допустить, что в мире не произойдёт ничего экстраординарного и прогнозируемое событие должно случиться в тех же условиях, в которых собирались экспериментальные данные, достаточно просто подставить в полученную функцию значение регрессора  $X$ . Например, при  $X = 22$  получим:

$$y = 0,53226 \times 22 - 0,91935 \approx 10,45.$$

Таким образом, используя метод наименьших квадратов, можно определить не только будущие значения переменной, но и те значения, которые она принимала ранее при условии, что  $X$  – величина, меняющаяся в зависимости от времени или же сама является показателем времени. Отталкиваясь от этой идеи, мы можем рассматривать настоящее состояние Вселенной как следствие его прошлого и причину его будущего. И пусть невозможно подвергнуть анализу все данные о силах, приводящих природу в движение, это и приводит к необходимости искать новые методы решения нерешаемых задач. Возможно, однажды науке удастся найти все ответы, и тогда уже не останется ничего неясного, а будущее – как и прошлое – предстанет настоящим в наших глазах.

### Библиографический список

1. Абдурахманова Н.Р. Уравнения, содержащие модуль // Ta'lim fidooyilari. – 2022. – №7.
2. Орлов А.И. Вероятностно-статистические модели корреляции и Регрессии // Научный журнал КубГАУ. – 2020. – №160.
3. Шукуров И.А. Парная регрессия и корреляция // Universum: технические науки. – 2021. – №5-1 (86).

## THE METHOD OF LEAST SQUARES IN A PROBLEM THAT DOES NOT HAVE A SOLUTION

**A.M. Zhukavin, Student**  
**Plekhanov Russian University of Economics**  
**(Russia, Moscow)**

***Abstract.** In this article, the author examines the potential of using the least squares method in order to predict the values of random variables that are in linear dependence. The author also reveals the inconsistency of the least absolute deviations method for solving problems of this kind and substantiates the idea that with a sufficient number of observations of two correlated variables, the law of causal relationship between them can be revealed as reliably as possible. For this purpose, in particular, an example of solving the problem of drawing a straight line through a cloud of points that do not lie on one straight line is given.*

***Keywords:** forecast, causal law, regressor, resulting value, linear dependence, deviation.*