УДК 519.65, 004.94

РЕШЕНИЕ ЗАДАЧИ ЛИНЕЙНОЙ РЕГРЕССИИ В СИСТЕМЕ MATHCAD

Дашкина А.Р., Бобылёв Ю.Ю.

Научный руководитель: Сабитова Г.С.

Аннотация. Задачи линейной регрессии относятся к большому кругу математических задач обработки экспериментальных данных. Применение компьютерных систем, в частности, MathCAD, дает возможность численного решения таких задач. Целью данной статьи является решение задачи регрессионного анализа методом наименьших квадратов с использованием инструментов математической системы MathCAD.

Ключевые слова: экспериментальные данные, задачи линейной регрессии, таблично заданная функция, система MathCAD.

SOLVING THE LINEAR REGRESSION PROBLEM IN MATHCAD

A. R. Dashkina, Y. Y. Bobylev

Scientific adviser: G. S. Sabitova

Annotation. Linear regression problems belong to a large range of mathematical problems for processing experimental data. The use of computer systems, in particular, MathCAD, makes it possible to solve such problems numerically. The purpose of this article is solving the problem of regression analysis by the least squares method using the tools of the mathematical system MathCAD.

Keywords: experimental data, linear regression problems, table-defined function, MathCAD system.

Интерполяционные формулы в точности воспроизводят значения заданной функции в узлах интерполяции [1]. Однако для ряда случаев выполнение этого условия затруднительно или даже нецелесообразно, например:

1. Если заданные величины x и y являются экспериментальными данными, то могут содержать в себе существенные ошибки. Поэтому построение аппроксимирующего многочлена, воспроизводящего в точности заданное значение функции, означало бы тщательное копирование допущенных при измерениях ошибок.

2. Если имеются точные значения функции в некоторых точках, но число таких точек n весьма велико, то интерполяционный многочлен будет очень высокой степени, что нежелательно для проведения интерполяции.

Поэтому возникает задача построения многочлена некоторой вполне определенной степени, но меньшей, чем $n - 1$, который, хотя и не дает точных значений функции в узлах интерполяции, достаточно близко к ним подходит.

Задачи математической регрессии имеют смысл приближения выборки данных (x_i, y_i) некоторой функцией $f(x)$, определенным образом минимизирующей совокупность ошибок $|f(x_i) - y_i|$. Регрессия сводится к подбору неизвестных коэффициентов, определяющих аналитическую зависимость $f(x)$ – некоторую приближенную к искомой функцию.

Для того, чтобы разобраться с применением алгоритмов математической обработки экспериментальных данных [2], последовательно выполним ряд заданий в программе MathCAD версии 14.0.0.163.

Задание 1. Создайте таблицу экспериментальных данных:

$x_i = a + h \cdot i$, где $h = (b - a)/10$, $i = 0,1, \dots ,10$ на отрезке $[a, b]$.

Решение. Этапы выполнения задания в программе MathCAD приведены на рис. 1.

$a := 2 \qquad b := 4$

$i := 0..10 \qquad h := \frac{(b-a)}{10} \qquad x_i := a + h \cdot i$

$x =$

	0
0	2
1	2.2
2	2.4
3	2.6
4	2.8
5	3
6	3.2
7	3.4
8	3.6
9	3.8
10	4

$$y := \begin{pmatrix} 4.70 \\ 4.64 \\ 4.57 \\ 4.45 \\ 4.40 \\ 4.34 \\ 4.27 \\ 4.37 \\ 4.42 \\ 4.50 \\ 4.62 \end{pmatrix}$$

Рис. 1. Создание таблицы экспериментальных данных

Задание 2. Аппроксимировать многочленами 2-ой и 6-ой степени по *методу наименьших квадратов* функцию, заданную таблицей значений x_i и y_i и сравнить качество приближений. Построить графики многочленов и отметить узловые точки (x_i, y_i).

Решение. Наиболее часто используемый и простой вид регрессии – *линейная*. Приближение данных (x_i, y_i) выполняется линейной функцией $y(x) = a \cdot x + b$. Также линейную регрессию часто называют методом наименьших квадратов, потому как коэффициенты a и b вычисляются из условия минимизации суммы квадратов ошибок $|a \cdot x_i + b - y_i|$. *Метод наименьших квадратов* (*МНК*) состоит в следующем: для данных значений $x = x_0, x_1, \dots, x_n$ и $y = y_0, y_1, \dots, y_n$ подобрать многочлен заданной степени $m < n$ вида [2]

$$\varphi(x) = a_m x^m + a_{m-1} x^{m-1} + \ldots + a_1 x + a_0, \qquad (1)$$

принимающий в заданных точках x_i значения как можно более близкие к табличным значениям y_i. Коэффициенты a_i многочлена (1) находят из решения системы

$$\begin{cases} b_{00}a_0 + b_{01}a_1 + \dots + b_{0m}a_m = c_0, \\ b_{10}a_0 + b_{11}a_1 + \dots + b_{1m}a_m = c_1, \\ \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ b_{m0}a_0 + b_{m1}a_1 + \dots + b_{mm}a_m = c_m \cdot \end{cases}, \quad (2)$$

где

$$b_{k,l} = \sum_{i=0}^{n} x_i^{k+l},\ c_k = \sum_{i=0}^{n} x_i^k y_i, \quad k,l = 0,1,\dots,m \;. \quad (3)$$

Проведем аппроксимацию многочленом 2-ой степени функции, заданной таблицей значений x_i, y_i, построим график многочлена и отметим узловые точки (x_i, y_i). Этапы выполнения аппроксимации представлены на рис. 2-4.

$m := 2$ - степень многочлена

$j := 0..m$

$A_{i,j} := (x_i)^j$

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 1 & 2.2 & 4.84 \\ 1 & 2.4 & 5.76 \\ 1 & 2.6 & 6.76 \\ 1 & 2.8 & 7.84 \\ 1 & 3 & 9 \\ 1 & 3.2 & 10.24 \\ 1 & 3.4 & 11.56 \\ 1 & 3.6 & 12.96 \\ 1 & 3.8 & 14.44 \\ 1 & 4 & 16 \end{pmatrix}$$

- матрица коэффициентов при неизвестных системы линейных уравнений

$Z := A^T \cdot A$

Рис. 2. Нахождение коэффициентов интерполяционного многочлена

$c := A^T \cdot y$ - вектор свободных членов системы линейных уравнений

$d := Z^{-1} \cdot c$

$$d = \begin{pmatrix} 7.543 \\ -2.053 \\ 0.329 \end{pmatrix}$$ - решение системы линейных алгебраических уравнений

$t := 0, 0.1 .. x_{last(x)}$

$\phi(x) := \sum_{i=0}^{m} \left(d_i \cdot x^i\right)$ - вычисление расчетных значений функции отклика

$\sum_{i=0}^{last(x)} \left(y_i - \phi(x_i)\right)^2 = 0.014$ - вычисление погрешности

Рис. 3. Аппроксимация таблично заданной функции многочленом 2-ой степени

Построим график многочлена $\varphi(x)$ и отметим узловые точки x_i и y_i.

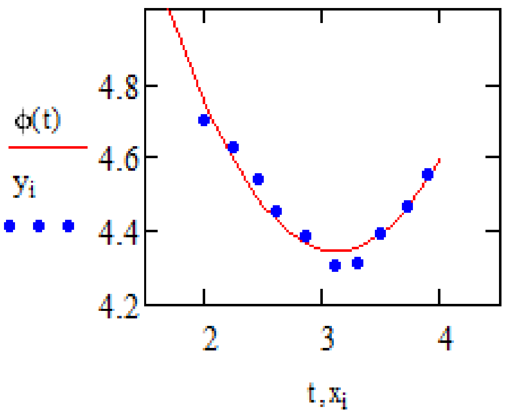

Рис. 4. Построение результативного графика

Проведем аналогичную процедуру для многочлена 6-ой степени. Этапы проведения аппроксимации представлены на рис. 5-7.

$m := 6$ - степень многочлена

$j := 0..m$

$A_{i,j} := (x_i)^j$

$$A = \begin{pmatrix} 1 & 2 & 4 & 8 & 16 & 32 & 64 \\ 1 & 2.2 & 4.84 & 10.648 & 23.426 & 51.536 & 113.38 \\ 1 & 2.4 & 5.76 & 13.824 & 33.178 & 79.626 & 191.103 \\ 1 & 2.6 & 6.76 & 17.576 & 45.698 & 118.814 & 308.916 \\ 1 & 2.8 & 7.84 & 21.952 & 61.466 & 172.104 & 481.89 \\ 1 & 3 & 9 & 27 & 81 & 243 & 729 \\ 1 & 3.2 & 10.24 & 32.768 & 104.858 & 335.544 & 1.074\times10^3 \\ 1 & 3.4 & 11.56 & 39.304 & 133.634 & 454.354 & 1.545\times10^3 \\ 1 & 3.6 & 12.96 & 46.656 & 167.962 & 604.662 & 2.177\times10^3 \\ 1 & 3.8 & 14.44 & 54.872 & 208.514 & 792.352 & 3.011\times10^3 \\ 1 & 4 & 16 & 64 & 256 & 1.024\times10^3 & 4.096\times10^3 \end{pmatrix}$$

- матрица коэффициентов при неизвестных системы линейных уравнений

$Z := A^T \cdot A$

Рис. 5. Аппроксимация функции многочленом 6-ой степени

$c := A^T \cdot y$ - вектор свободных членов системы линейных уравнений

$d := Z^{-1} \cdot c$

$$d = \begin{pmatrix} 131.5 \\ -274.222 \\ 243.253 \\ -113.006 \\ 28.954 \\ -3.882 \\ 0.213 \end{pmatrix}$$

- решение системы линейных алгебраических уравнений

$t := 0, 0.1 .. x_{last(x)}$

$\phi(x) := \sum_{i=0}^{m} \left(d_i \cdot x^i\right)$ - вычисление расчетных значений функции отклика

$\sum_{i=0}^{last(x)} \left(y_i - \phi(x_i)\right)^2 = 3.385 \times 10^{-3}$ - вычисление погрешности

Рис. 6. Аппроксимация функции многочленом 6-ой степени

Рис. 7. Построение результативного графика

Задание 3. Для приведенных в таблице экспериментальных данных (x_i, y_i) определить параметры *линейной регрессии* с использованием встроенных функций MathCAD *slope* и *intercept*. Отобразить графически совокупность точек векторов x_i и y_i и результаты проведенной линейной регрессии. Сравнить также с дублирующим способом, использующим функцию *line*.

Решение. Перед началом выполнения задания, разберемся с понятием «регрессионный анализ».

Пусть имеются два ряда чисел $x = x_0, x_1, \dots, x_n$ и $y = y_0, y_1, \dots, y_n$, при этом предполагается, что ряд y каким-либо образом зависит от ряда x. *Задача регрессионного анализа* состоит в восстановлении математической зависимости (*регрессии*) $y(x)$ исходя из экспериментальных данных (x_i, y_i), где $i = 0,1, \dots n$.

MathCAD включает ряд функций для вычисления регрессии [1]. Функции отличаются прежде всего типом используемой ими для аппроксимации данных кривой.

Для расчета линейной регрессии в MathCAD присутствует два дублирующих друг друга способа: с помощью функции *line* и функций *intercept* и *slope*.

Встроенные функции *line*, *intercept* (to intercept – отложить отрезок на линии) и *slope* (наклон) решают самую простую и распространенную задачу *линейной регрессии* экспериментальных данных:

$\varphi(x)$ = *line(vx, vy);*

$\varphi(x) = slope(vx, vy)\ x + intercept(vx, vy)$

Решение задачи посредствам пакета MathCAD приведено на рис. 8-9.

$t := 2, 2.1 .. x_{10}$

$sl := slope(x, y)$ - наклон линии для данных из x и y

$interc := intercept(x, y)$ - смещение по оси ординат линии регрессии для данных из x и y

$\varphi(t) := sl \cdot t + interc$ - линия регрессии

Отобразим графически совокупность точек векторов x_i и y_i и результат проведенной линейной регрессии при помощи встроенных функций slope и intercept.

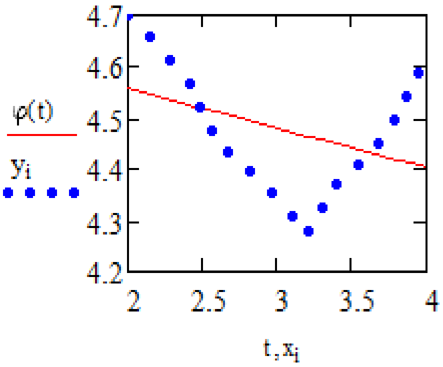

Рис. 8. Определение параметров линейной регрессии с использованием встроенных функций MathCAD *slope* и *intercept*

$\varphi(t) := line(x, y)_0 + line(x, y)_1 \cdot t$ - линия регрессии

Отобразим графически совокупность точек векторов x_i и y_i и результат проведенной линейной регрессии при помощи встроенной функции line.

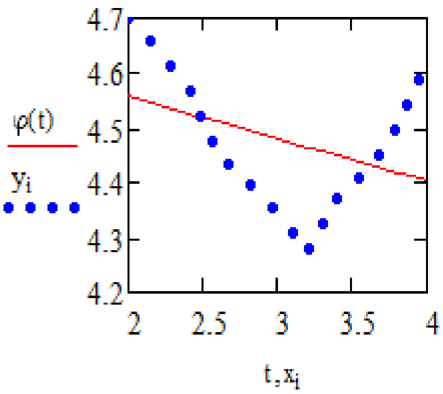

Рис. 9. Определение параметров линейной регрессии с использованием встроенной функции MathCAD *line*

Как мы можем заметить, результаты дублирующих способов линейной регрессии идентичны.

Заключение. В статье рассмотрены задачи математической обработки экспериментальных данных, разрешаемые при помощи метода наименьших квадратов и линейного регрессионного анализа, когда функция задана таблично. Приведено решение задач с помощью инструментов математической системы MathCAD.

Данная работа будет полезной для подготовки магистрантов направления «Математическое моделирование» во время самостоятельной работы [3], [4].

Список литературы

1. Дашкина, А. Р., Бобылев, Ю. Ю. Интерполяция и предсказание функций в MathCAD / А. Р. Дашкина, Ю. Ю. Бобылев // Kazakhstan Science Journal. –2019. – № 1 (14). – Т. 3. – С. 13-24.
2. Сабитова, Г. С., Калиев, И. А. Практикум по информационным технологиям. Часть 1: учеб. пособие / Г. С. Сабитова, И. А. Калиев. – Стерлитамак: РИО Стерлитамакского филиала БашГУ. – 2015. – 127с.
3. Сабитова, Г. С., Калиев, И. А., Суйундукова, А. К. Разработка дистанционного учебного курса «Пакеты математического моделирования» / Г. С. Сабитова, И. А. Калиев, А. К. Суйундукова // Kazakhstan Science Journal. –2019. – № 1 (2). – Т. 2. – С. 5-12.
4. Сабитова, Г. С., Калиев, И. А., Суйундукова, А. К. Создание компьютерного практикума «Пакеты математического моделирования» / Г. С. Сабитова, И. А. Калиев, А. К. Суйундукова // Kazakhstan Science Journal. –2019. – № 4 (5). – Т. 2. – С. 5-12.

Сведения об авторах

Дашкина Алина Рустемовна – студент факультета математики и информационных технологий, Стерлитамакский филиал Башкирского государственного университета; Россия, Стерлитамак.

Бобылев Юрий Юрьевич – студент факультета математики и информационных технологий, Стерлитамакский филиал Башкирского государственного университета; Россия, Стерлитамак.

Сабитова Гульнара Сагындыковна – кандидат физико-математических наук, доцент кафедры прикладной информатики и программирования, Стерлитамакский филиал Башкирского государственного университета; Россия, г. Стерлитамак.

About authors

Dashkina Alina Rustemonva – Student of Faculty of Mathematics and Information Technology, Sterlitamak branch of Bashkir State University; Russia, Sterlitamak.

Bobylev Yuri Yurievich – Student of Faculty of Mathematics and Information Technology, Sterlitamak branch of Bashkir State University; Russia, Sterlitamak.

Sabitova Gulnara Sagyndykovna – Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of Applied Informatics and Programming, Sterlitamak branch of the Bashkir State University; Russia, Sterlitamak.