

Московский государственный технический университет  
имени Н. Э. Баумана



Факультет: Информатика и системы управления

Кафедра: Программное обеспечение ЭВМ и информационные технологии

Дисциплина: Математическая статистика

**Лабораторная работа №3**  
**Метод наименьших квадратов**

Выполнил:  
Группа:  
Вариант: 15

Москва, 2015 г.

# Содержание

<b>1</b>	<b>Постановка задачи</b>	<b>3</b>
<b>2</b>	<b>Теоретическая часть</b>	<b>4</b>
2.1	Аппроксимация неизвестной зависимости . . . . .	4
2.2	Понятие МНК-оценки параметров линейной модели . . . . .	4
2.3	Формулы для вычисления МНК-оценки в рассматриваемом случае . . . .	5
2.4	Среднеквадратичное отклонение . . . . .	5
<b>3</b>	<b>Текст программы</b>	<b>5</b>
<b>4</b>	<b>Результаты расчётов</b>	<b>6</b>

# 1 Постановка задачи

**Цель работы:** аппроксимация неизвестной зависимости параболой.

**Содержание работы:**

1. Для выборки  $(y_i, t_i)$ ,  $i = \overline{1; n}$ , реализовать в виде программы на ЭВМ:
  - (а) вычисление МНК-оценки  $\vec{\theta} = (\theta_0, \theta_1, \theta_2)$  параметров модели  $y = \theta_0 + \theta_1 t + \theta_2 t^2$ ;
  - (б) вычисление среднеквадратичного отклонения  $\Delta = \sqrt{\sum_{i=1}^n (y_i - y(t_i))^2}$  полученной модели от результатов наблюдений;
  - (в) построение на одном графике системы точек  $(y_i, t_i)$ ,  $i = \overline{1; n}$ , и графика функции  $y = y(t)$ ,  $t \in [t_{(1)}; t_{(n)}]$  (для полученной оценки  $\vec{\theta}$ ).
2. провести необходимые вычисления и построить соответствующие графики для выборки из индивидуального варианта.

**Содержание отчёта:**

1. постановка задачи аппроксимации неизвестной зависимости по результатам наблюдений;
2. понятие МНК-оценки параметров линейной модели;
3. формулы для вычисления МНК-оценки в рассматриваемом случае;
4. текст программы;
5. результаты расчетов и графики для выборки из индивидуального варианта.

## 2 Теоретическая часть

### 2.1 Аппроксимация неизвестной зависимости

Предположим, что в нашем распоряжении имеются результаты  $n$  наблюдений:

$$\begin{cases} y_1 = \Phi(x_1) + \varepsilon_1 \\ \dots \\ y_n = \Phi(x_n) + \varepsilon_n \end{cases}, \quad \text{где} \quad (1)$$

- $y_1, \dots, y_n$  —  $n$  реализаций  $Y$ ;
- $\varepsilon_1, \dots, \varepsilon_n$  —  $n$  реализаций  $\varepsilon$ ;
- $x_1, \dots, x_n$  — известные значения.

Требуется на основе этих данных подобрать функцию  $\hat{\Phi}$  так, чтобы она наилучшим образом аппроксимировала неизвестную функцию  $\Phi$ .

### 2.2 Понятие МНК-оценки параметров линейной модели

Часто в качестве функции  $\hat{\Phi}(x)$  выбирают функцию следующего вида:

$$\hat{\Phi}(x) = \theta_1 \psi_1(x) + \dots + \theta_p \psi_p(x), \quad \text{где} \quad (2)$$

- $\psi_1, \dots, \psi_p$  — известные функции.

Параметры  $\theta_1, \dots, \theta_p$  подбирают так, чтобы  $\hat{\Phi}(x)$  наилучшим образом аппроксимировала  $\Phi(x)$ .

С учётом предположения о виде функции  $\hat{\Phi}$  результаты наблюдений можно записать в виде:

$$y_i = \theta_1 \psi_1(x_i) + \dots + \theta_p \psi_p(x_i) + \varepsilon_i, \quad i = \overline{1; n}. \quad (3)$$

В матричном виде:

$$\vec{y} = \Psi \vec{\theta} + \vec{\varepsilon}, \quad \text{где} \quad (4)$$

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \Psi = \begin{pmatrix} \psi_1(x_1) & \psi_2(x_1) & \dots & \psi_p(x_1) \\ \psi_1(x_2) & \psi_2(x_2) & \dots & \psi_p(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(x_n) & \psi_2(x_n) & \dots & \psi_p(x_n) \end{pmatrix}, \quad \vec{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}, \quad \vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Задача заключается в подборе  $\vec{\theta}$ .

Будем предполагать, что:

1.  $M\varepsilon = 0$ , т. е. систематические ошибки отсутствуют;
2.  $\varepsilon \sim N(0, \sigma^2)$ .

**Определение 2.1.** Оценка  $\hat{\vec{\theta}}$  вектора  $\vec{\theta}$  называется *оценкой*, полученной по *методу наименьших квадратов (МНК-оценкой)*, если  $\hat{\vec{\theta}}$  доставляет минимальное значение функции  $S(\vec{\theta}) = \|y - \Psi \vec{\theta}\|^2$ .

## 2.3 Формулы для вычисления МНК-оценки в рассматриваемом случае

В рассматриваемом случае МНК-оценка вектора  $\vec{\theta}$  имеет вид:

$$\hat{\vec{\theta}} = (\Psi^T \Psi)^{-1} \cdot \Psi^T \vec{y}, \quad \text{где} \quad (5)$$

- $\text{rg}(\Psi) = p$  — числу столбцов

причём так как  $y = \theta_0 + \theta_1 t + \theta_2 t^2$ , то

$$\Psi = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{pmatrix}. \quad (6)$$

## 2.4 Среднеквадратичное отклонение

Среднеквадратичное отклонение полученной модели от результатов наблюдений будем вычислять как

$$\Delta = \sqrt{\sum_{i=1}^n (y_i - y(t_i))^2}, \quad \text{где} \quad (7)$$

## 3 Текст программы

```
1 function lw3()
2     t0Array = csvread('t.csv');
3     y0Array = csvread('y.csv');
4
5     y1Array = OLS(y0Array, t0Array);
6
7     delta = calculateDelta(y0Array, y1Array);
8     fprintf('Delta = %3.2f', delta);
9
10    plot(t0Array, y0Array, '.b', t0Array, y1Array, 'r');
11    legend({
12        'Original sample';
13        'Output model';
14    });
15 end
16
17 function Y = OLS(yArray, tArray)
18     psiMatrix = makePsiMatrix(tArray);
19
20     thetaArray = (psiMatrix' * psiMatrix) \ (psiMatrix' * yArray);
21     disp(thetaArray);
22
23     n = length(yArray);
```

```

24     Y = zeros(1, n);
25     for i = 1:n
26         Y(i) = thetaArray(1) + ...
27               thetaArray(2)*tArray(i) + ...
28               thetaArray(3)*tArray(i)*tArray(i);
29     end
30 end
31
32 function Psi = makePsiMatrix(tArray)
33     n = length(tArray);
34     p = 3;
35     Psi = zeros(n, p);
36     for i = 1:n
37         Psi(i, 1) = 1;
38         Psi(i, 2) = tArray(i);
39         Psi(i, 3) = tArray(i)*tArray(i);
40     end
41 end
42
43 function X = calculateDelta(yArray, yNewArray)
44     n = length(yArray);
45     sum = 0;
46     for i = 1:n
47         sum = sum + power((yArray(i) - yNewArray(i)), 2);
48     end
49     X = sqrt(sum);
50 end

```

## 4 Результаты расчётов

*МНК-оценка* в данном варианте имеет вид

$$\hat{\theta} = \begin{pmatrix} 5.0505 \\ 2.4455 \\ -5.2084 \end{pmatrix}. \quad (8)$$

В свою очередь *среднеквадратичного отклонение* полученной модели от результатов наблюдений равна

$$\Delta = 170.21. \quad (9)$$

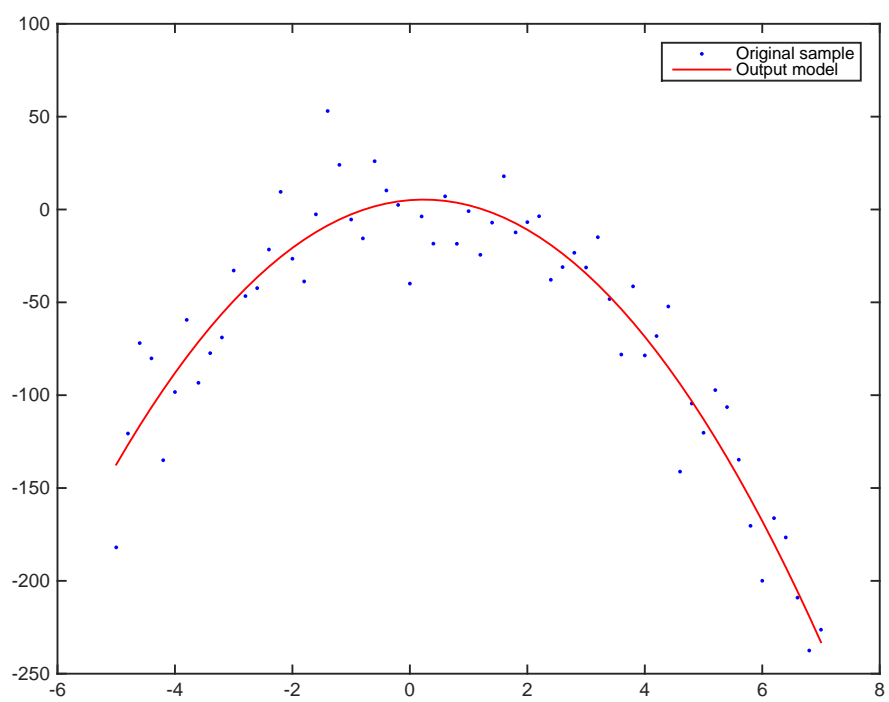


Рис. 1: График исходной выборки и полученной модели