Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

Отчет по лабораторной работе №3 по дисциплине «Математическая статистика»

Тема Метод наименьших квадратов

Студент Криков А.В.

Группа ИУ7-63Б

Оценка (баллы) ____________________

Преподаватель Власов П. А.

Москва — 2022 г.

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 $\Delta = \sqrt{\sum_{i=1}^{n} \left(y_i - y(t_i)\right)^2}$ полученной модели от результатов наблюдений;

   (в) построение на одном графике системы точек (y_i, t_i), $i = \overline{1; n}$, и графика функции $y = y(t)$, $t \in \left[t_{(1)}; t_{(n)}\right]$ (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 Постановка задачи аппроксимации неизвестной зависимости по результатам наблюдений

Пусть Y — случайная величина, $X_1, \dots, X_s$ — детерминированные величины. Если изменение значений $X_1, \dots, X_s$ влияет на значения случайной величины Y, то говорят, что Y стохастически зависит от $X_1, \dots, X_s$. Задача регрессионного анализа — задача, связанная с установлением аналитических зависимостей между случайной величиной Y и детерминированными величинами $X_1, \dots, X_s$, носящими количественный характер. В регрессионном анализе используется модель черного ящика, как наиболее общая модель, ассоциируемая с понятием отображения. На вход поступает вектор $X_1, \dots, X_s$, который посредством некоторого отображения Φ и случайных возмущений $\varepsilon_1, \dots, \varepsilon_n$ преобразуется в вектор $Y_1, \dots, Y_m$.

2.2 Понятие МНК-оценки параметров линейной модели

Предположим, что в нашем распоряжении имеются результаты n наблюдений:

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

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

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

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

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

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

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

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

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

В матричном виде:

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

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

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 $\vec{\theta}$ 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:

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

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

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

Среднеквадратичное отклонение полученной модели от результатов наблюдений будем вычислять как

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

- y_i — результат наблюдения
- $y(t_i)$ — результат аппроксимации

3 Текст программы

```
% Вариант 11
function lab_3()
    clc
    T = [-5.00,-4.80,-4.60,-4.40,-4.20,-4.00,-3.80,-3.60,-3.40,-3.20, ...
        -3.00,-2.80,-2.60,-2.40,-2.20,-2.00,-1.80,-1.60,-1.40,-1.20, ...
        -1.00,-0.80,-0.60,-0.40,-0.20,0.00,0.20,0.40,0.60,0.80,1.00, ...
        1.20,1.40,1.60,1.80,2.00,2.20,2.40,2.60,2.80,3.00,3.20,3.40, ...
        3.60,3.80,4.00,4.20,4.40,4.60,4.80,5.00,5.20,5.40,5.60,5.80, ...
        6.00,6.20,6.40,6.60,6.80,7.00];
    Y = [-23.33,-51.17,-9.60,-25.13,35.42,-29.66,-16.56,-26.60,-9.52, ...
        -55.24,-26.38,-9.29,-30.18,-13.03,-2.67,-44.04,-22.10,14.26, ...
        1.90,15.23,-10.97,16.15,41.97,2.41,14.54,47.54,-1.73,19.05, ...
        -16.56,19.20,-13.71,-7.97,-35.79,4.26,12.11,8.49,-22.74, ...
        13.50,-40.73,-21.16,-19.63,10.61,-6.27,-31.38,-10.91,-9.06, ...
        -27.11,-15.97,-12.55,-37.36,-22.32,-12.61,-21.23,-10.29, ...
        -28.08,  -40.09,-33.55,1.42,-4.48,-30.29,-62.62];

```

```
    Yt = OLS(Y, T);

    delta = sqrt(sum((Y - Yt) .^ 2));
    fprintf('Delta = %.4f\n', delta);

    plot(T, Y, '.b', T, Yt, 'r');
    legend({
        'Система точек (t_i,y_i)';
        'Полученная модель';
    });
end

function Yt = OLS(Y, T)
    T2 = T .* T;
    psi = horzcat(ones(length(T), 1), T', T2');

    theta = (psi' * psi) \ (psi' * Y');
    fprintf("theta1 = %.4f\n", theta(1));
    fprintf("theta2 = %.4f\n", theta(2));
    fprintf("theta3 = %.4f\n", theta(3));
    fprintf("Yt = %.4f + %.4f*t + %.4f*t^2\n", theta(1), theta(2), theta(3));

    Yt = theta(1) + theta(2) * T + theta(3) * T2;
end
```

4 Результаты расчётов

МНК-оценка в данном варианте имеет вид

$$\hat{\vec{\theta}} = \begin{pmatrix} -0.7795 \\ 0.9662 \\ -0.8992 \end{pmatrix}. \tag{8}$$

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

$$\Delta = 154.2021\,. \tag{9}$$

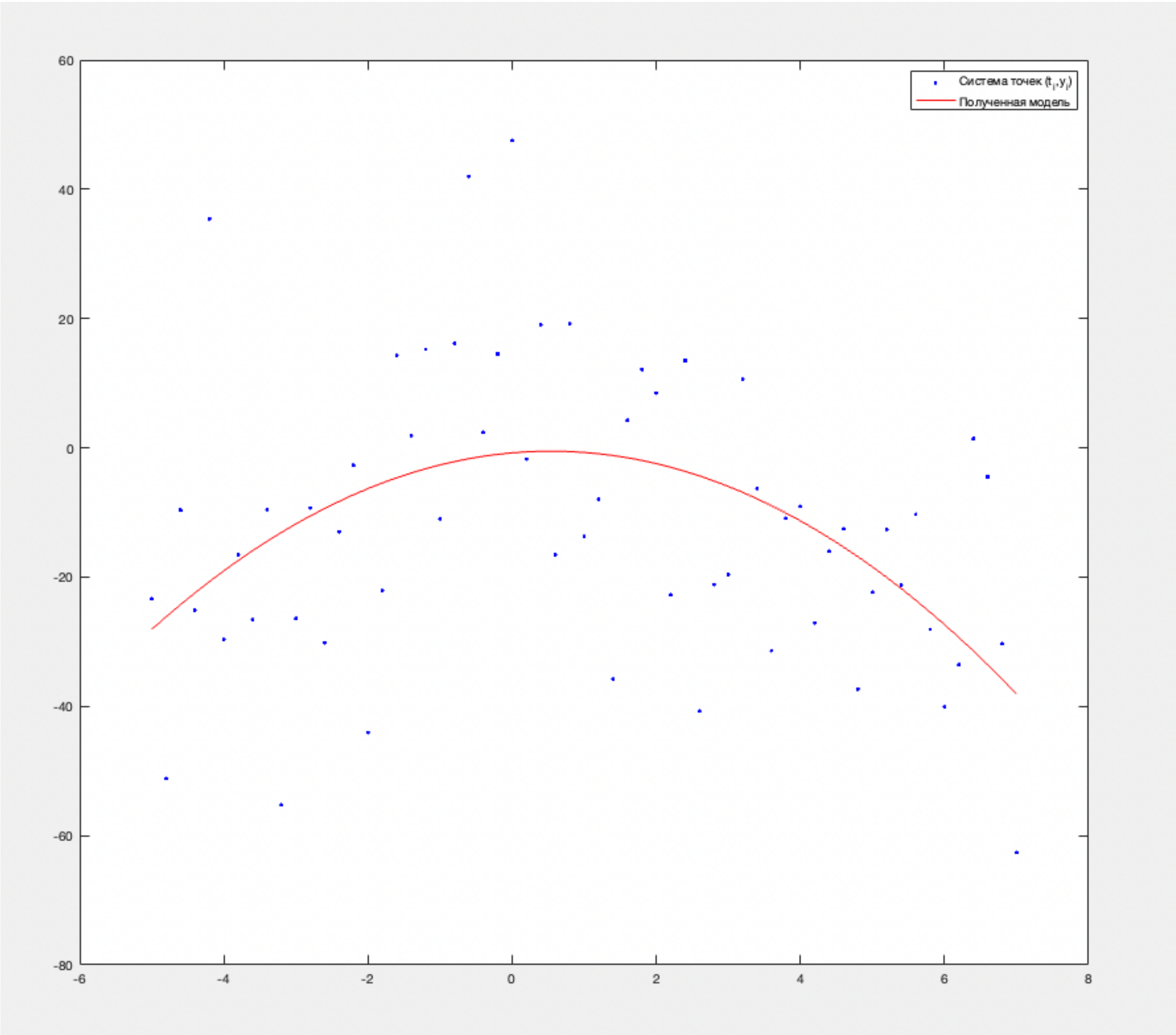

Рис. 1: График исходной выборки и полученной модели