МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ

Московский авиационный институт

(национальный исследовательский университет)

Институт № 8

Информационных технологий и прикладной математики

Кафедра 804 «Теория вероятностей и компьютерное моделирование»

КУРСОВАЯ РАБОТА

по дисциплине «Теория вероятностей и математическая статистика»

на тему:«Метод наименьших квадратов»

Выполнил: студент группы М8О-303Б-20

(Фамилия, имя, отчество)

(подпись)

Принял: доцент кафедры 804

Игнатов Алексей Николаевич

(Фамилия, имя, отчество)

(подпись)

Оценка: **Дата**:

Москва, 2022

Содержание

Задание:

Модель полезного сигнала имеет вид:

$$y(x) = \theta_0 + \theta_1 x + ... + \theta_m x^m. \qquad (1)$$

Рассматривается модель наблюдений:

$$y_k = \theta_0 + \theta_1 x_k + ... + \theta_m x_k^m + \varepsilon_k; \quad k = \overline{1, n}. \qquad (2)$$

где $\varepsilon_1, \ldots, \varepsilon_n$ – независимые и одинаково распределённые случайные величины.

Смоделировать два набора наблюдений на основе модели (2) для следующих случаев:

1 случай	2 случай
$m = 3$, $\varepsilon_k \sim \mathcal{N}(0, \sigma^2)$	$m = 2$, $\varepsilon_k \sim \mathcal{R}(-3\sigma, 3\sigma)$
$x_k = -4 + k \cdot \frac{8}{n}$, $k = \overline{1, n}$, $n = 40$.	

Для обоих случаев выполнить по очереди следующие задания:

1. Подобрать порядок многочлена $\hat{m}$ в модели (1), используя критерий Фишера, и вычислить оценки неизвестных параметров $(\theta_0, \ldots, \theta_{\hat{m}})$ методом наименьших квадратов.

2. В предположении нормальности ошибок построить доверительные интервалы уровней надёжности α_1=0.95 и α_2=0.99 для параметров $(\theta_0, \ldots, \theta_{\hat{m}})$.

3. В предположении нормальности ошибок построить доверительные интервалы уровней надёжности α_1=0.95 и α_2=0.99 для полезного сигнала (1).

4. Представить графически:

   - Истинный полезный сигнал.
   - набор наблюдений.
   - Оценку полезного сигнала, полученную в шаге 1.
   - Доверительные интервалы полезного сигнала, полученные в шаге 3.

5. По остаткам регрессии построить оценку плотности распределения случайной ошибки наблюдения в виде гистограммы.

6. Вычислить оценку дисперсии σ^2 случайной ошибки.

7. По остаткам регрессии с помощью χ^2-критерия Пирсона на уровне значимости 0.05 проверить гипотезу о том, что закон распределения ошибки наблюдения является нормальным.

Вариант задания — 20: $\theta_0 = 20, \theta_1 = -2, \theta_2 = -1, \theta_3 = -0.06, \sigma^2 = 2.6$

Решение задачи будет реализовано на языке программирования Python.

Перед тем, как приступать к непосредственному решению задачи необходимо сгенерировать выборку объёма 40 для каждого из случаев. Сделаем это спомощью функций random.normal() и random.uniform() из пакета numpy:

$$E_1 = \begin{pmatrix} -0.15942282 \\ -0.8410887 \\ \dots \\ -2.414894773 \end{pmatrix} \qquad E_2 = \begin{pmatrix} 3.133377073 \\ -0,826984262 \\ \dots \\ 1.310205996 \end{pmatrix}$$

1 Номер 1

Для подбора старшей степени многочлена в модели (1) будем использовать критерий Фишера. Основная и альтернативная гипотеза этого критерия имеют вид:

$$H_0 : \theta_m = 0; \quad H_A : \theta_m \neq 0 \ ,$$

Статистика критерия имеет вид:

$$Z = \frac{\hat{\theta}_p^2}{\frac{\alpha_{p+1}}{n-(p+1)}(Y - X\hat{\Theta})^T(Y - X\hat{\Theta})} \ ,$$

где n — объём выборки,$Y_{n\times 1}$ — выборка, $\hat{\Theta}$ — матрица МНК-оценок параметров θ, α_{p+1} — p+1-ый элемент главной диагонали $(Y - X\hat{\Theta})^T(Y - X\hat{\Theta})$.

$$X_{n\times(p+1)} = \begin{pmatrix} f_1(x_{11}, \ldots, x_{1m}) & \cdots & f_{p+1}(x_{11}, \ldots, x_{1m}) \\ \vdots & \ddots & \vdots \\ f_1(x_{n1}, \ldots, x_{nm}) & \cdots & f_{p+1}(x_{n1}, \ldots, x_{nm}) \end{pmatrix}$$

$$Y = \begin{pmatrix} 16.29289718 \\ 16.1982713 \\ \cdots \\ -10.25489477 \end{pmatrix}$$

Будем считать точечные МНК-оценки для каждого порядка по очереди, начиная с 1.

1.1 p = 1:

$$X_{40\times 2} = \begin{pmatrix} 1 & -3.8 \\ 1 & -3.6 \\ \dots & \dots \\ 1 & 4 \end{pmatrix} \qquad (X^TX)^{-1}_{2\times 2} = \begin{pmatrix} 0.0250469 & -0.00046904 \\ -0.00046904 & 0.00469043 \end{pmatrix}$$

$$\hat{\Theta}_{2\times 1} = (X^TX)^{-1}X^TY \approx \begin{pmatrix} 14.93460792 \\ -2.75534714 \end{pmatrix}$$

$$(Y - X\hat{\Theta})^T(Y - X\hat{\Theta}) \approx 1092.3899849648342$$

$$Z \approx 56.19021242056394$$

Уровени надёжности будем считать за 0,95. Следовательно, $f_{0.95,1,38} \approx 4.098$.

$$Z > f_{0.95,1,38}$$

Статистика попала в критическую область. Следовательно, гипотеза отклоняется.

1.2 p = 2:

$$X_{40\times 3} = \begin{pmatrix} 1 & -3.8 & 14.44 \\ 1 & -3.6 & 12.96 \\ \dots & \dots & \dots \\ 1 & 4 & 16 \end{pmatrix} (X^TX)^{-1}_{3\times 3} = \begin{pmatrix} 0.05623827 & 0.00070356 & -0.00586304 \\ 0.00070356 & 0.00473451 & -0.00022042 \\ -0.00586304 & -0.00022042 & 0.00110208 \end{pmatrix}$$

$$\hat{\Theta}_{3\times 1} = (X^TX)^{-1}X^TY \approx \begin{pmatrix} 20.52748388 \\ -2.54508865 \\ -1.05129247 \end{pmatrix}$$

$$(Y - X\hat{\Theta})^T(Y - X\hat{\Theta}) \approx 89.54009617325156$$

$$Z \approx 414.40033539268376 \qquad f_{0.95,1,37} \approx 4.1055$$

$$Z > f_{0.95,1,37}$$

Статистика попала в критическую область. Следовательно, гипотеза отклоняется.

1.3 p = 3:

$$X_{40\times 4} = \begin{pmatrix} 1 & -3.8 & 14.44 & -54.872 \\ 1 & -3.6 & 12.96 & -46.656 \\ \dots & \dots & \dots & \dots \\ 1 & 4 & 16 & 64 \end{pmatrix}$$

$$(X^T X)^{-1}_{4\times 4} = \begin{pmatrix} 5.65 & -1.76 & -5.94 & 2.58 \\ -1.76 & 2.93 & 5.52 & -2.57 \\ -5.94 & 5.52 & 1.13 & -8.08 \\ 2.58 & -2.57 & -8.08 & 2.69 \end{pmatrix}$$

$$\hat{\Theta}_{4\times 1} = (X^T X)^{-1} X^T Y \approx \begin{pmatrix} 20.49060076 \\ -2.1770278 \\ -1.03973761 \\ -0.0385162 \end{pmatrix}$$

$$(Y - X\hat{\Theta})^T (Y - X\hat{\Theta}) \approx 84.033028275468$$

$$Z \approx 2.359244316059978 \qquad f_{0.95,1,36} \approx 4.113189609999999$$

$$Z < f_{0.95,1,36}$$

Статистика не попала в критическую область. Следовательно, гипотеза принимается. Это означает, что для оценки сигнала требуется порядок многочлена модели $\hat{m} = 2$.

1.4 Результат работы программы:

```
Модель имеет порядок 2
[20.52748388 -2.54508865 -1.05129247]
```

Программа вывела интересующий нас ответ и вектор-строку коэффициентов $\hat{\theta}_k$.

2 Номер 2

В предположении нормальности ошибок для построения доверительных интервалов параметров θ используется следующая формула:

$$\hat{\theta}_k - t_{1-\frac{\alpha}{2},n-(p+1)}\frac{||\hat{E}||\sqrt{\alpha_k}}{\sqrt{n-(p+1)}} \leqslant \theta_k \leqslant \hat{\theta}_k + t_{1-\frac{\alpha}{2},n-(p+1)}\frac{||\hat{E}||\sqrt{\alpha_k}}{\sqrt{n-(p+1)}},$$

где α_k — k-й элемент главной диагонали матрицы $(Y - X\hat{\Theta})^T(Y - X\hat{\Theta})$, а α — уровень значимости.

Следовательно, для уровня надёжности $= 0.99$:

$$19.52573442 \leqslant \theta_0 \leqslant 21.52923334$$

$$-2.83574569 \leqslant \theta_1 \leqslant -2.2544316$$

$$-1.19152499 \leqslant \theta_2 \leqslant -0.91105996$$

2.1 Результат работы программы:

```
[[19.52573442 21.52923334]
 [-2.83574569 -2.2544316 ]
 [-1.19152499 -0.91105996]]
```

3 Номер 3

В предположении нормальности ошибок для построения доверительных интервалов полезного сигнала $\varphi(x,\theta) = \theta_0 + \theta_1 x + \ldots + \theta_{\hat{m}} x^{\hat{m}}$ используется следующая формула:

$$\varphi(x,\hat{\theta}) - t_{1-\frac{\alpha}{2},n-(p+1)} \frac{||\hat{E}||\sqrt{\alpha(x)}}{\sqrt{n-(p+1)}} \leqslant \varphi(x,\theta) \leqslant \varphi(x,\hat{\theta}) + t_{1-\frac{\alpha}{2},n-(p+1)} \frac{||\hat{E}||\sqrt{\alpha(x)}}{\sqrt{n-(p+1)}},$$

где $\alpha(x) = (1, x, \ldots, x^{\hat{m}})(X^T X)^{-1}(1, x, \ldots, x^{\hat{m}})^T$

Следовательно, для уровня надёжности $= 0.99$ при $\hat{m} = 2$:

$$14.54478336 \leqslant \varphi(x_0,\theta) \leqslant 18.35985664$$

$$15.31707454 \leqslant \varphi(x_1,\theta) \leqslant 18.76164546$$

$$\ldots$$

3.1 Результат работы программы:

```
[[14.54478336 18.35985664]
 [15.31707454 18.76164546]
 [16.04484046 19.15163954]
 [16.72413412 19.52802588]
 [17.35089158 19.88910842]
 [17.92104058 20.23319942]
 [18.43071065 20.55840935]
 [18.87654945 20.86233055]
 [19.25608023 21.14167977]
 [19.5679512  21.3920488 ]
 [19.81191671 21.60792329]
 [19.98851385 21.78300615]
 [20.09857214 21.91070786]
 [20.14275995 21.98460005]
 [20.12129399 21.99870601]
 [20.033822   21.947618  ]
 [19.87942392 21.82649608]
 [19.65667004 21.63100996]
 [19.36369294 21.35726706]
 [18.99825054 21.00174946]
 [18.55777054 20.56126946]
 [18.03937294 20.03294706]
 [17.43987004 19.41420996]
 [16.75574392 18.70281608]
 [15.983102   17.896898  ]
 [15.11761399 16.99502601]
 [14.15443995 15.99628005]
 [13.08817214 14.90030786]
 [11.91283385 13.70732615]
 [10.62199671 12.41800329]
 [ 9.2090712  11.0331688 ]
 [ 7.66776023  9.55335977]
 [ 5.99254945  7.97833055]
 [ 4.17903065  6.30672935]
 [ 2.22392058  4.53607942]
 [ 0.12481158  2.66302842]
 [-2.12018588  0.68370588]
 [-4.51275954 -1.40596046]
 [-7.05460546 -3.61003454]
 [-9.74753664 -5.93246336]]
```

4 Номер 4

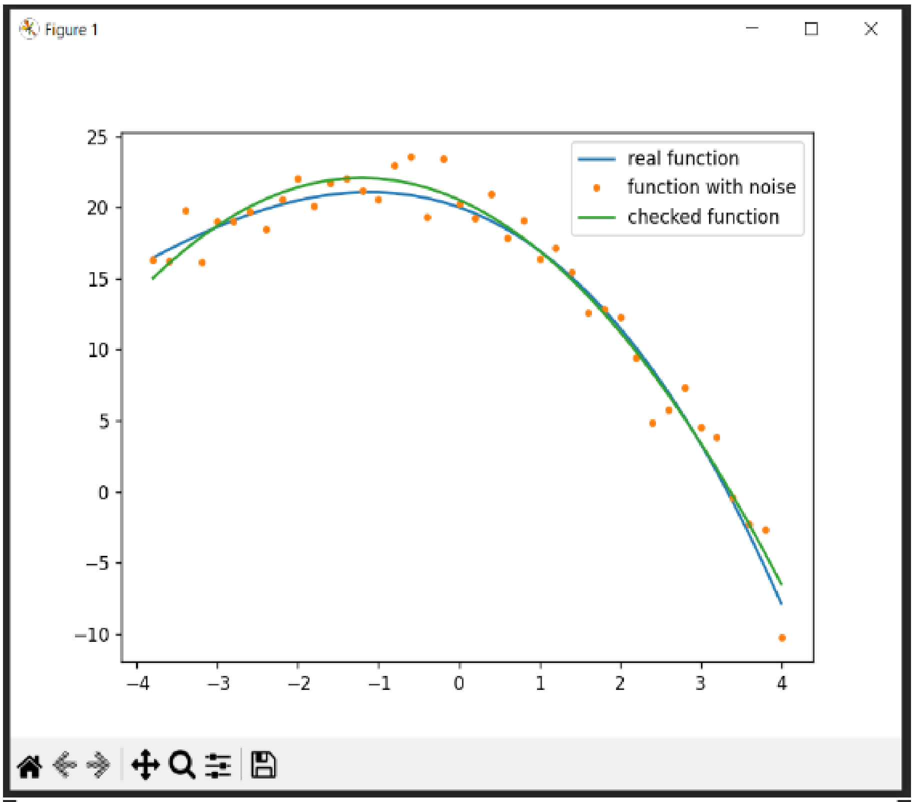


Figure 1
real function
function with noise
checked function
25
20
15
10
5
0
−5
−10
−4
−3
−2
−1
0
1
2
3
4

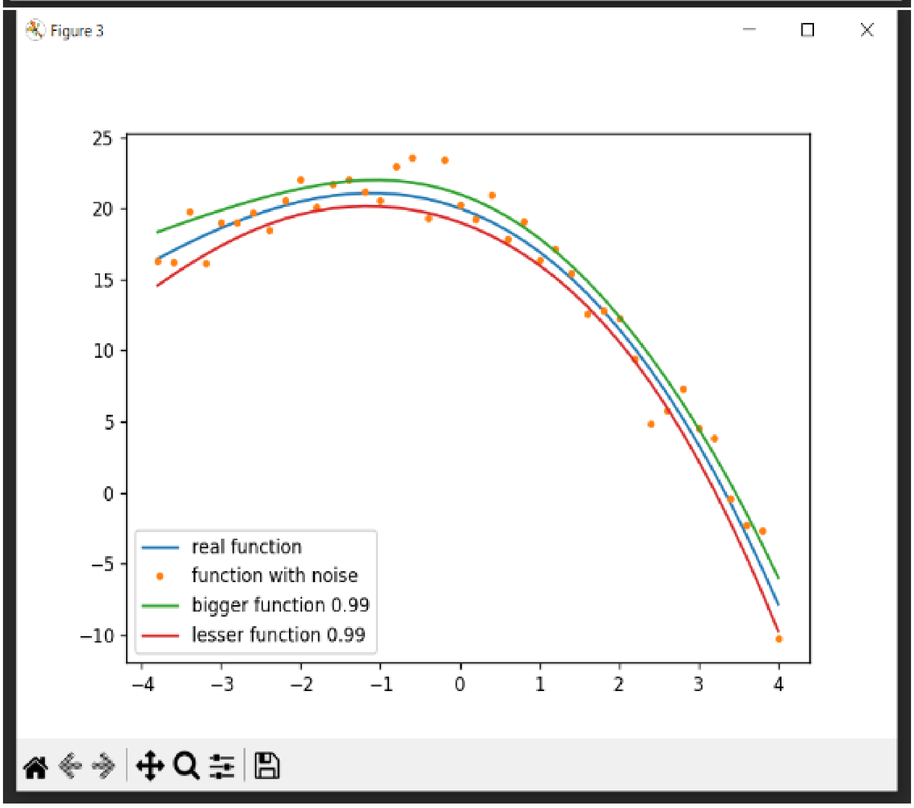


Figure 3
real function
function with noise
bigger function 0.99
lesser function 0.99
25
20
15
10
5
0
−5
−10
−4
−3
−2
−1
0
1
2
3
4

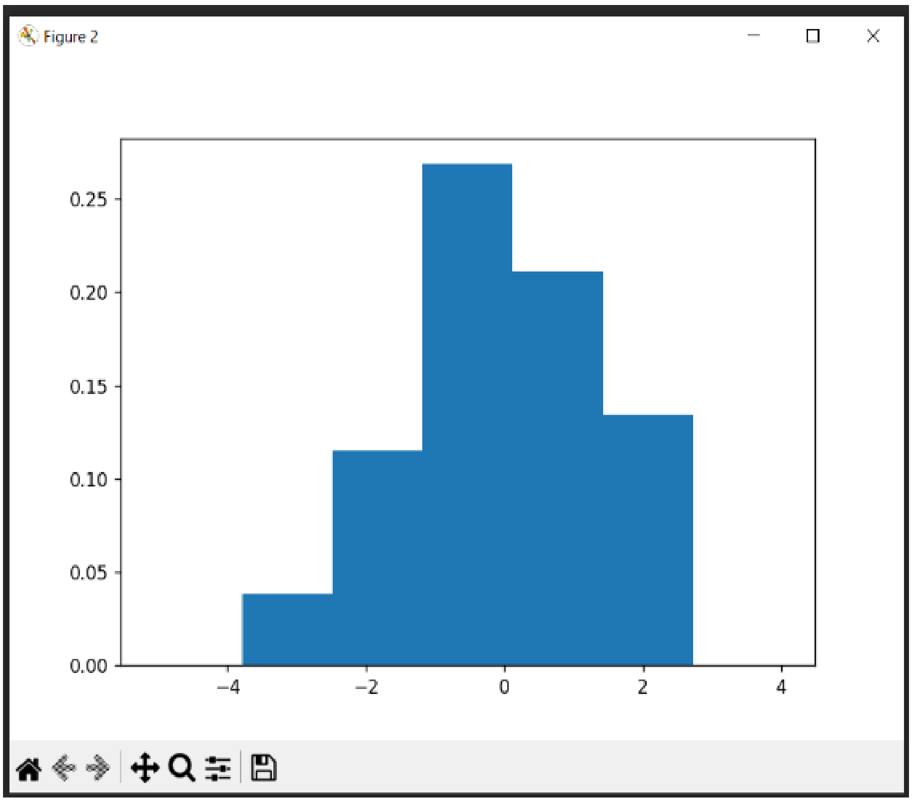
5 Номер 5

Гистограмма является оценкой плотности совместного распределения выборки. Общий её вид:

$$\hat{f}_n(x) = \begin{cases} \frac{n_k}{n(t_{k+1}-t_k)} & x \in [t_k;\ t_{k+1}) \\ 0 & x \in [t_0;\ t_1) \cup [t_l;\ t_{l+1}) \end{cases} \tag{1}$$

Для нашего случая положим $l = 6$. Тогда:

$$\hat{f}_n(x) = \begin{cases} 0 & x \in (-\infty;\ -3.781) \\ \frac{2}{40 \cdot 1.301} & x \in [-3.781;\ -2.48 \\ \frac{6}{40 \cdot 1.301} & x \in [-2.48;\ -1.179) \\ \frac{14}{40 \cdot 1.301} & x \in [-1.179;\ 0.122) \\ \frac{11}{40 \cdot 1.301} & x \in [0.122;\ 1.423) \\ \frac{7}{40 \cdot 1.301} & x \in [1.423;\ 2.724] \\ 0 & x \in (2.724;\ \infty) \end{cases} \tag{2}$$



6 Номер 6

Для рассчета несмещённой оценки дисперсии ошибок используется формула:

$$\hat{\sigma}^2 = \frac{||\hat{E}||^2}{n-(p+1)} = \frac{89.54009617325156}{37} = 2.420002599277069$$

7 Номер 7

Для проверки нормальности распределения ошибок требуется использовать χ^2 – критерий Пирсона. В данном случае основная и альтернативная гипотезы этого критерия имеют вид:

$$H_0 : E \sim \mathcal{N}(0,\theta); \quad H_A : E \nsim \mathcal{N}(0,\theta)$$

$$L = \prod_{k=1}^{40} \frac{1}{\sqrt{2\pi\theta}} \cdot exp\left\{-\frac{x_i^2}{2\theta}\right\} = (2\pi\theta)^{-20} \cdot exp\left\{-\frac{1}{2\theta}\sum_{k=1}^{40} x_i^2\right\}$$

$$\bar{L} = -20 \cdot ln(2\pi\theta) + \frac{1}{2\theta}\sum_{k=1}^{40} x_i^2$$

$$\frac{\partial \bar{L}}{\partial \theta} = -40\theta + \sum_{k=1}^{40} x_i^2 = 0$$

$$\theta = \frac{1}{40}\sum_{k=1}^{40} x_i^2 = \frac{1}{40}||\hat{E}||^2 = \frac{89.54009617325156}{40} = 2.238502404331289$$

Статистика $T(Z_n)$ считается по формуле:

$$T(Z_n) = n\sum_{k=0}^{l} \frac{(p_k - \hat{p}_k)^2}{p_k} \quad ,$$

где $p_k = \Phi_0\left(\frac{t_{k+1}-m}{\sqrt{\theta}}\right) - \Phi_0\left(\frac{t_k - m}{\sqrt{\theta}}\right)$, а $\hat{p}_k = \frac{n_k}{n}$.

В нашем случае $T(Z_n) = 2.6005633514085758$. Рассчитаем квантиль $\chi^2_{0{,}95}(5) \approx 11.0705$

$$0 < T(Z_n) = 2.6005633514085758 < \chi^2_{0{,}95}(5) \approx 11.0705$$

Статистика попала в доверительный интервал. Следовательно, гипотеза H_0 — принимается, и закон распределения ошибок действительно окажется нормальным.

```
Распределение ошибок нормальное
```

Случай 2:

8 Номер 1

Для подбора старшей степени многочлена в модели (1) будем использовать критерий Фишера. Основная и альтернативная гипотеза этого критерия имеют вид:

$$H_0 : \theta_m = 0; \quad H_A : \theta_m \neq 0 \ ,$$

Статистика критерия имеет вид:

$$Z = \frac{\hat{\theta}_p^2}{\frac{\alpha_{p+1}}{n-(p+1)}(Y - X\hat{\Theta})^T(Y - X\hat{\Theta})} \ ,$$

где n — объём выборки,$Y_{n\times 1}$ — выборка, $\hat{\Theta}$ — матрица МНК-оценок параметров θ, α_{p+1} — p+1-ый элемент главной диагонали $(Y - X\hat{\Theta})^T(Y - X\hat{\Theta})$.

$$X_{n\times(p+1)} = \begin{pmatrix} f_1(x_{11},\dots,x_{1m}) & \cdots & f_{p+1}(x_{11},\dots,x_{1m}) \\ \vdots & \ddots & \vdots \\ f_1(x_{n1},\dots,x_{nm}) & \cdots & f_{p+1}(x_{n1},\dots,x_{nm}) \end{pmatrix}$$

$$Y = \begin{pmatrix} 11.69022427 \\ 10.35798324 \\ \dots \\ -6.38975253 \end{pmatrix}$$

Будем считать точечные МНК-оценки для каждого порядка по очереди, начиная с 1.

8.1 p = 1:

$$X_{40\times 2} = \begin{pmatrix} 1 & -3.8 \\ 1 & -3.6 \\ \dots & \dots \\ 1 & 4 \end{pmatrix} \qquad (X^T X)^{-1}_{2\times 2} = \begin{pmatrix} 0.0250469 & -0.00046904 \\ -0.00046904 & 0.00469043 \end{pmatrix}$$

$$\hat{\Theta}_{2\times 1} = (X^T X)^{-1} X^T Y \approx \begin{pmatrix} 14.26809312 \\ -2.0637206 \end{pmatrix}$$

$$(Y - X\hat{\Theta})^T (Y - X\hat{\Theta}) \approx 1433.8385376061742$$

$$Z \approx 24.064250951693666$$

Уровени надёжности будем считать за 0,95. Следовательно, $f_{0.95,1,38} \approx 4.098$.

$$Z > f_{0.95,1,38}$$

Статистика попала в критическую область. Следовательно, гипотеза отклоняется.

8.2 p = 2:

$$X_{40\times 3} = \begin{pmatrix} 1 & -3.8 & 14.44 \\ 1 & -3.6 & 12.96 \\ \dots & \dots & \dots \\ 1 & 4 & 16 \end{pmatrix} (X^T X)^{-1}_{3\times 3} = \begin{pmatrix} 0.05623827 & 0.00070356 & -0.00586304 \\ 0.00070356 & 0.00473451 & -0.00022042 \\ -0.00586304 & -0.00022042 & 0.00110208 \end{pmatrix}$$

$$\hat{\Theta}_{3\times 1} = (X^T X)^{-1} X^T Y \approx \begin{pmatrix} 20.27721818 \\ -1.83781365 \\ -1.12953479 \end{pmatrix}$$

$$(Y - X\hat{\Theta})^T (Y - X\hat{\Theta}) \approx 276.15984600437747$$

$$Z \approx 155.1062263721988 \qquad f_{0.95,1,37} \approx 4.1055$$

$$Z > f_{0.95,1,37}$$

Статистика попала в критическую область. Следовательно, гипотеза отклоняется.

8.3 p = 3:

$$X_{40\times 4} = \begin{pmatrix} 1 & -3.8 & 14.44 & -54.872 \\ 1 & -3.6 & 12.96 & -46.656 \\ \dots & \dots & \dots & \dots \\ 1 & 4 & 16 & 64 \end{pmatrix}$$

$$(X^T X)^{-1}_{4\times 4} = \begin{pmatrix} 5.65 & -1.76 & -5.94 & 2.58 \\ -1.76 & 2.93 & 5.52 & -2.57 \\ -5.94 & 5.52 & 1.13 & -8.08 \\ 2.58 & -2.57 & -8.08 & 2.69 \end{pmatrix}$$

$$\hat{\Theta}_{4\times 1} = (X^T X)^{-1} X^T Y \approx \begin{pmatrix} 2.02910317 \\ -1.97566032 \\ -1.13386233 \\ 1.44251438 \end{pmatrix}$$

$$(Y - X\hat{\Theta})^T (Y - X\hat{\Theta}) \approx 275.3873899714874$$

$$Z \approx 0.1009792684658641 \qquad f_{0.95,1,36} \approx 4.113189609999999$$

$$Z < f_{0.95,1,36}$$

Статистика не попала в критическую область. Следовательно, гипотеза принимается. Это означает, что для оценки сигнала требуется порядок многочлена модели $\hat{m} = 2$.

8.4 Результат работы программы:

```
Модель имеет порядок 2
[20.27721818 -1.83781365 -1.12953479]
```

Программа вывела интересующий нас ответ и вектор-строку коэффициентов $\hat{\theta}_k$.

9 Номер 2

В предположении нормальности ошибок для построения доверительных интервалов параметров θ используется следующая формула:

$$\hat{\theta}_k - t_{1-\frac{\alpha}{2},n-(p+1)} \frac{||\hat{E}||\sqrt{\alpha_k}}{\sqrt{n-(p+1)}} \leqslant \theta_k \leqslant \hat{\theta}_k + t_{1-\frac{\alpha}{2},n-(p+1)} \frac{||\hat{E}||\sqrt{\alpha_k}}{\sqrt{n-(p+1)}},$$

где α_k — k-й элемент главной диагонали матрицы $(Y - X\hat{\Theta})^T(Y - X\hat{\Theta})$, а α — уровень значимости.

Следовательно, для уровня надёжности $= 0.99$:

$$18.51795585 \leqslant \theta_0 \leqslant 22.03648051$$

$$-2.34826262 \leqslant \theta_1 \leqslant -1.32736467$$

$$-1.37580971 \leqslant \theta_2 \leqslant -0.88325986$$

9.1 Результат работы программы:

```
[[18.51795585 22.03648051]
 [-2.34826262 -1.32736467]
 [-1.37580971 -0.88325986]]
```

10 Номер 3

В предположении нормальности ошибок для построения доверительных интервалов полезного сигнала $\varphi(x,\theta) = \theta_0 + \theta_1 x + \ldots + \theta_{\hat{m}} x^{\hat{m}}$ используется следующая формула:

$$\varphi(x,\hat{\theta}) - t_{1-\frac{\alpha}{2},n-(p+1)} \frac{||\hat{E}||\sqrt{\alpha(x)}}{\sqrt{n-(p+1)}} \leqslant \varphi(x,\theta) \leqslant \varphi(x,\hat{\theta}) + t_{1-\frac{\alpha}{2},n-(p+1)} \frac{||\hat{E}||\sqrt{\alpha(x)}}{\sqrt{n-(p+1)}},$$

где $\alpha(x) = (1, x, \ldots, x^{\hat{m}})(X^T X)^{-1}(1, x, \ldots, x^{\hat{m}})^T$

Следовательно, для уровня надёжности $= 0.99$ при $\hat{m} = 2$:

$$9.81000333 \leqslant \varphi(x_0,\theta) \leqslant 16.50999667$$

$$11.21533959 \leqslant \varphi(x_1,\theta) \leqslant 17.26466041$$

$$\ldots$$

10.1 Результат работы программы:

```
[[ 9.81000333 16.50999667] [17.80073767 21.31926233]
 [11.21533959 17.26466041] [17.28945258 20.79054742]
 [12.51193534 17.96806466] [16.70634204 20.17365796]
 [13.69791673 18.62208327] [16.05028572 19.46971428]
 [14.77120454 19.22879546] [15.31950535 18.68049465]
 [15.72970494 19.79029506] [14.51145395 17.80854605]
 [16.57167846 20.30832154] [13.62268946 16.85731054]
 [17.29629559 20.78370441] [12.64877272 15.83122728]
 [17.90426451 21.21573549] [11.58426534 14.73573466]
 [18.39826906 21.60173094] [10.42293565 13.57706435]
 [18.78293565 21.93706435] [ 9.15826906 12.36173094]
 [19.06426534 22.21573466] [ 7.78426451 11.09573549]
 [19.24877272 22.43122728] [ 6.29629559  9.78370441]
 [19.34268946 22.57731054] [ 4.69167846  8.42832154]
 [19.35145395 22.64854605] [ 2.96970494  7.03029506]
 [19.27950535 22.64049465] [ 1.13120454  5.58879546]
 [19.13028572 22.54971428] [-0.82208327  4.10208327]
 [18.90634204 22.37365796] [-2.88806466  2.56806466]
 [18.60945258 22.11054742] [-5.06466041  0.98466041]
 [18.24073767 21.75926233] [-7.34999667 -0.65000333]]
```

11 Номер 4

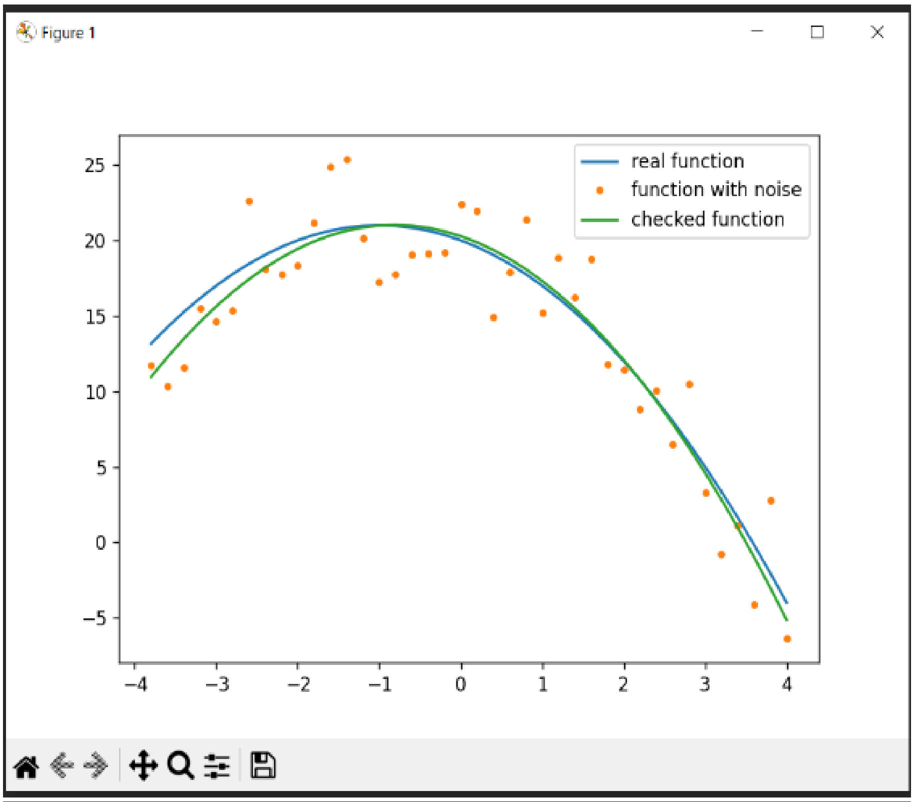


Figure 1
real function
function with noise
checked function
25
20
15
10
5
0
−5
−4
−3
−2
−1
0
1
2
3
4

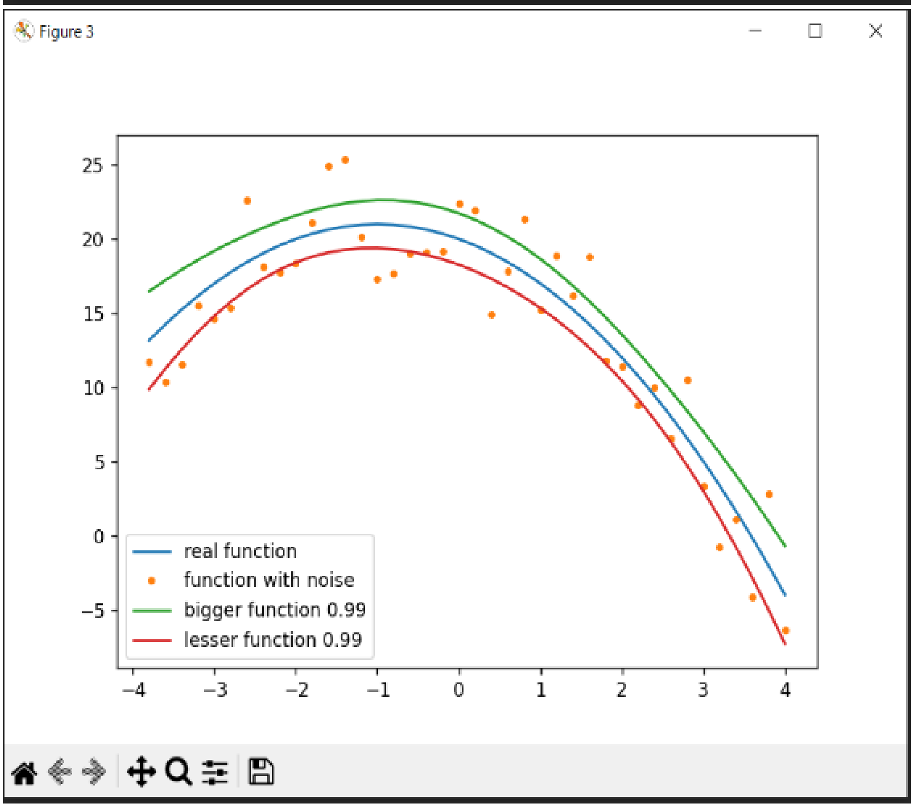


Figure 3
real function
function with noise
bigger function 0.99
lesser function 0.99
25
20
15
10
5
0
−5
−4
−3
−2
−1
0
1
2
3
4

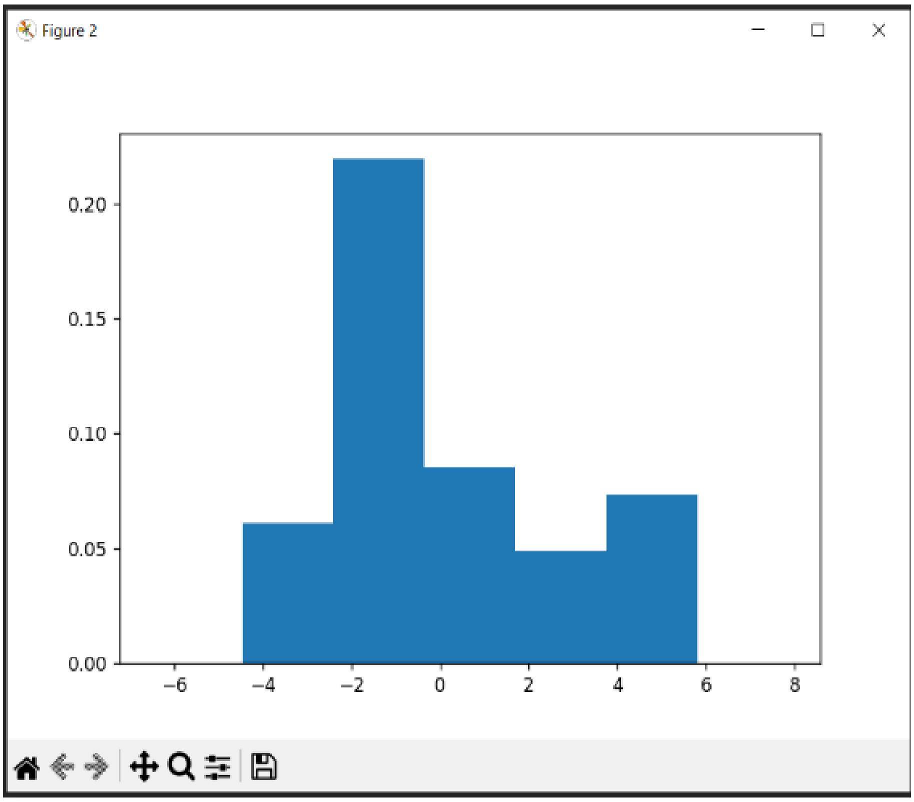
12 Номер 5

Гистограмма является оценкой плотности совместного распределения выборки. Общий её вид:

$$\hat{f}_n(x) = \begin{cases} \frac{n_k}{n(t_{k+1}-t_k)} & x \in [t_k;\ t_{k+1}) \\ 0 & x \in [t_0;\ t_1) \cup [t_l;\ t_{l+1}) \end{cases} \tag{3}$$

Для нашего случая положим $l = 6$. Тогда:

$$\hat{f}_n(x) = \begin{cases} 0 & x \in (-\infty;\ -4.454) \\ \frac{2}{40\cdot 2.05} & x \in [-4.454;\ -2.404) \\ \frac{6}{40\cdot 2.05} & x \in [-2.404;\ -0.354) \\ \frac{14}{40\cdot 2.05} & x \in [-0.354;\ 1.696) \\ \frac{11}{40\cdot 2.05} & x \in [1.696;\ 3.746) \\ \frac{7}{40\cdot 2.05} & x \in [3.746;\ 5.796) \\ 0 & x \in (5.796;\ \infty) \end{cases} \tag{4}$$



13 Номер 6

Для рассчета несмещённой оценки дисперсии ошибок используется формула:

$$\hat{\sigma}^2 = \frac{||\hat{E}||^2}{n-(p+1)} = \frac{276.1598460043775}{37} = 7.463779621739933$$

14 Номер 7

Для проверки нормальности распределения ошибок требуется использовать χ^2 – критерий Пирсона. В данном случае основная и альтернативная гипотезы этого критерия имеют вид:

$$H_0 : E \sim \mathcal{N}(0,\theta); \quad H_A : E \nsim \mathcal{N}(0,\theta)$$

$$L = \prod_{k=1}^{40} \frac{1}{\sqrt{2\pi\theta}} \cdot exp\left\{-\frac{x_i^2}{2\theta}\right\} = (2\pi\theta)^{-20} \cdot exp\left\{-\frac{1}{2\theta}\sum_{k=1}^{40} x_i^2\right\}$$

$$\bar{L} = -20 \cdot ln(2\pi\theta) + \frac{1}{2\theta}\sum_{k=1}^{40} x_i^2$$

$$\frac{\partial \bar{L}}{\partial \theta} = -40\theta + \sum_{k=1}^{40} x_i^2 = 0$$

$$\theta = \frac{1}{40}\sum_{k=1}^{40} x_i^2 = \frac{1}{40}||\hat{E}||^2 = \frac{276.1598460043775}{40} = 6.903996150109438$$

Статистика $T(Z_n)$ считается по формуле:

$$T(Z_n) = n\sum_{k=0}^{l} \frac{(p_k - \hat{p}_k)^2}{p_k} \quad ,$$

где $p_k = \Phi_0\left(\frac{t_{k+1}-m}{\sqrt{\theta}}\right) - \Phi_0\left(\frac{t_k - m}{\sqrt{\theta}}\right)$, а $\hat{p}_k = \frac{n_k}{n}$.

В нашем случае $T(Z_n) = 15.505909406453956$. Рассчитаем квантиль $\chi^2_{0,95}(5) \approx 11.0705$

$$0 < \chi^2_{0,95}(5) \approx 11.0705 < T(Z_n) = 15.505909406453956$$

Статистика не попала в доверительный интервал. Следовательно, гипотеза H_0 — отвергается, и закон распределения ошибок не является нормальным с параметрами $(0, \theta)$.

```
Не является нормальным
```

15 Вывод:

В ходе курсовой работы был применён метод наименьших квадратов для двух случаев. Мы убедились, что метод МНК-оценки хорошо работает независимо от того, какое распределение имеет вектор ошибок. Даже в случае с равномерным распределением, когда правомерность наших действий была под вопросом мы получили весьма точный ответ.