Розрахункова робота

з математичної статистики

студентки групи КА-92

Катасанової Карини

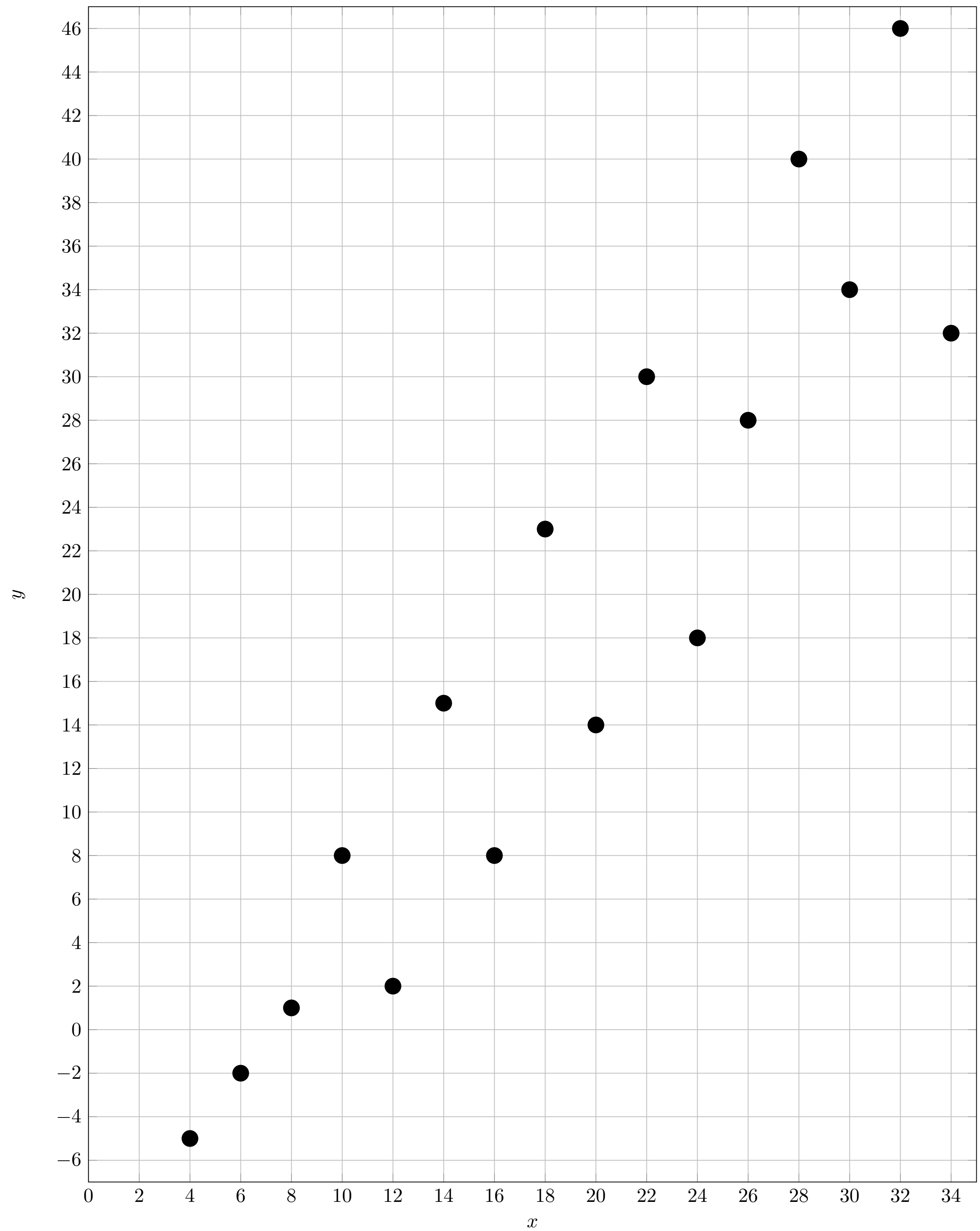
Завдання 1:
1.Провести аналіз вибірки та вибрати підходящу лінійну регресійну модель.
2.За методом найменших квадратів знайти оцінки параметрів вибраної моделі.
3.На рівні значущості $\alpha = 0.05$ перевірити адекватність побудованої моделі.
4.Для самого малого значення параметра побудованої моделі на рівні значущості $\alpha = 0.05$ перевірити гіпотезу про його значущість.
5.Побудувати прогнозований довірчий інтервал з довірчою ймовірністю $\gamma = 0.95$ для середнього значення відклику та самого значення відклику в точці $x = 16$.
6.Написати висновки.

Завдання відповідно до варіанту 7:

x_i	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
y_i	-5	-2	1	8	2	15	8	23	14	30	18	28	40	34	46	32

1.Провести аналіз вибірки та вибрати підходящу лінійну регресійну модель.

Задача, яка поставлена перед нами: дослідити зв'язок між відкликом Y та фактором X на основі вибірки $(x_i, y_i), i = \overline{1, 16}$.
Бачимо, що значення відклика залежить лише від одного фактора, обсяг вибірки $n = 16$.
Побудуємо графічне зображення вибірки:


46
44
42
40
38
36
34
32
30
28
26
24
22
20
18
16
14
12
10
8
6
4
2
0
−2
−4
−6
y
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34
x

Зважаючи на графічне зображення, можна припустити, що точки лежать (або є рівновіддаленими) на деякій прямій. Оберемо наступну модель:

$$y \;=\; f(x) \;=\; \beta_0 + \beta_1 x$$

2.За методом найменших квадратів знайти оцінки параметрів вибраної моделі.

Позначимо $\varepsilon_i = \eta_i - E\left(\eta/\xi = x_i\right)$ - це похибки спостережень, що дорівнюють різниці між отриманим значенням та усередненим значенням η при певному значенні ξ. Про розподіл $\overrightarrow{\varepsilon} = \left(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n\right)^T$, як правило, мало, що відомо: припускається, що його координати незалежні, однакого розподілені, їх дисперсія невідома, але математичне сподівання нульове:

$$E\varepsilon_i = E\Big(\eta_i - E\left(\eta/\xi = x_i\right)\Big) =$$

$$= E\eta_i - E\Big(E\left(\eta/\xi = x_i\right)\Big) = E\left(\eta/\xi = x_i\right) - E\left(\eta/\xi = x_i\right) = 0$$

Небхідно за значеннями $x_1, ...x_n$ та $\eta_1, ..., \eta_n$ оцінити $f(x)$. При цьому величини $x_1, ...x_n$ вважаємо невипадковими, а вся "випадковість" зосереджена в $\eta_1, ..., \eta_n$ та $\varepsilon_1, ..., \varepsilon_n$.
Оскільки $\eta_i = f(x_i) + \varepsilon_i$, то розподіл η_i буде таким самим як і розподіл ε_i, але з математичним сподіванням $f(x_i)$.
Припустимо, що $\overrightarrow{\varepsilon} \sim N(0; \sigma^2\mathbb{I})$, де $\mathbb{I}$ - одинична матриця розміру $n \times n$. Тоді:

$$\mathcal{L}\left(\eta_1, ..., \eta_n, \overrightarrow{\beta}\right) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}\left(\eta_k - f(x_k)\right)^2\right\}$$

При фіксованому значенні σ максимум функції правдоподібності по $\overrightarrow{\beta}$ досягається при наймешному значенні $\sum_{k=1}^{n}\left(\eta_k - f(x_k)\right)^2 = \sum_{k=1}^{n}\varepsilon_k^2$. Цей окремий випадок ММП називається методом найменших квадратів.
Оцінкою методу найменших квадратів для параметра $\overrightarrow{\beta} = (\beta_1, ..., \beta m)^T$ називається такий набір параметрів, при яких досягається мінімум $\sum_{k=1}^{n}\varepsilon_k^2$.
ОМНК для $\overrightarrow{\beta}$:

$$\overrightarrow{\beta^*} = \left(F^T F\right) F^T \overrightarrow{\eta} = A^{-1} F^T \overrightarrow{\eta}$$

Тут: A - інформаційна матриця, A^{-1} - дисперсійна матриця Фішера, F - матриця плану, $\overrightarrow{\eta}$ - відклик.
Усі підрахунки матриць та норм, що наведені у роботі були виконані за допомогою написаної мною програми на Python.
Запишемо матрицю плану:

$$F = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 & 20 & 22 & 24 & 26 & 28 & 30 & 32 & 34 \end{pmatrix}^T$$

Тепер запишемо інформаційну матрицю:

$$A = F^{\mathrm{T}} F = \begin{pmatrix} 16 & 304 \\ 304 & 7136 \end{pmatrix}$$

Дисперсійна матриця:

$$A^{-1} = \begin{pmatrix} \frac{223}{680} & -\frac{19}{1360} \\ -\frac{19}{1360} & \frac{1}{1360} \end{pmatrix} \approx \begin{pmatrix} 0.32794 & -0.01397 \\ -0.01397 & 0.00074 \end{pmatrix}$$

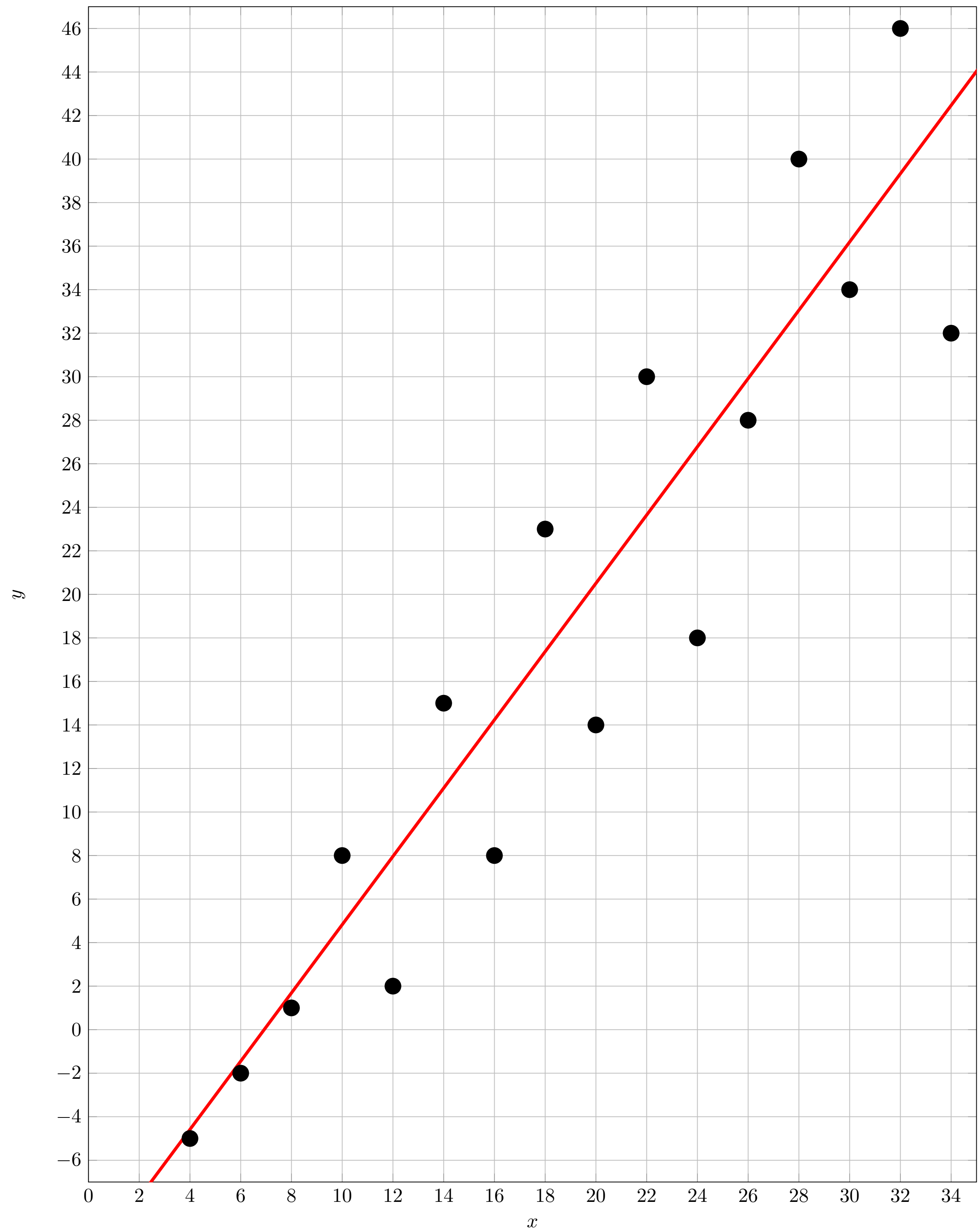
Вектор значень відкликів:

$$\overrightarrow{\eta_{\text{зн}}} = \begin{pmatrix} -5 & -2 & 1 & 8 & 2 & 15 & 8 & 23 & 14 & 30 & 18 & 28 & 40 & 34 & 46 & 32 \end{pmatrix}^T$$

Знайдемо значення ОМНК для $\overrightarrow{\beta}$:

$$\overrightarrow{\beta^*_{\text{зн}}} = A^{-1} F^T \overrightarrow{\eta_{\text{зн}}} = \begin{pmatrix} -10.86056 \\ 1.56844 \end{pmatrix}$$

Отже, маємо модель: $y = f^*(x) = -10.86056 + 1.56844x$. Зобразимо її графік на рисунку:


y
x
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34
−6 −4 −2 0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46

3. На рівні значущості $\alpha = 0.05$ перевірити адекватність побудованої моделі.

Для перевірки моделі на адекватність скористаємось F - критерієм: порівняння залишкової оцінки дисперсії $\left(\sigma^2\right)^{**}$, що обумовлена випадковими помилками вимірів та, можливо, неврахованими факторами, з незміщеною оцінкою дисперсії $D^{**}\eta$. Оскільки:

$$\frac{n-m}{\sigma^2}\left(\sigma^2\right)^{**} = \frac{n}{\sigma^2}\left(\sigma^2\right)^{*} = \frac{1}{\sigma^2}\|\overrightarrow{\eta} - F\overrightarrow{\beta^*}\|^2 \sim \chi^2_{n-m}$$

$$\frac{n-1}{\sigma^2}D^{**}\eta = \frac{n}{\sigma^2}D^{*}\eta = \frac{1}{\sigma^2}\sum_{k=1}^{n}\left(\eta_k - \overline{\eta}\right)^2 \sim \chi^2_{n-1}$$

Тоді:

$$\zeta = \frac{D^{**}\eta}{\left(\sigma^2\right)^{**}} \sim F(n-1, n-m)$$

Тут: n - обсяг вибірки, m - кількість параметрів моделі.
Оскільки F - критерій перевіряє чи є побудована модель кращою за найпростішу - константну, то:

$H_0:$ Константа і побудована моделі не відрізняються (тобто дисперсії їх похибок однакові)

$H_1:$ Побудована модель є кращою за константну

Критична область при цьому є правосторонньою.
Статистика, яка відповідає виправленій вибірковій дисперсії для η:

$$(D^{**}\eta) = \frac{1}{n-1}\sum_{k=1}^{n}(\eta_k - \overline{\eta})^2$$

$\overline{\eta}$- вибіркове середнє. Статистика, яка відповідає вибірковому середньому:

$$\overline{\eta} = \frac{1}{n}\sum_{k=1}^{n}\eta_k$$

Обчислимо значення вибіркового середнього:

$$\overline{y} = \frac{1}{16}(-5-2+1+8+2+15+8+23+14+30+18+28+40+34+46+32) = \frac{292}{16} = 18.25$$

Тепер можемо обчислити значення виправленої вибіркової дисперсії для η:

$$(D^{**}\eta)_{\text{зн}} = \frac{1}{15}\sum_{k=1}^{16}(y_k - \overline{y})^2 =$$

$$= \frac{(-5-18.25)^2 + (-2-18.25)^2 + (1-18.25)^2 + (8-18.25)^2 + (2-18.25)^2}{15} +$$

$$+\frac{(15-18.25)^2+(8-18.25)^2+(23-18.25)^2+(14-18.25)^2+(30-18.25)^2+(18-18.25)^2}{15}+$$

$$+\frac{(28-18.25)^2+(40-18.25)^2+(34-18.25)^2+(46-18.25)^2+(32-18.25)^2}{15}=$$

$$=\frac{540.5625+410.0625+297.5625+105.0625+264.0625+10.5625+105.0625+22.5625}{15}+$$

$$+\frac{18.0625+138.0625+0.0625+95.0625+473.0625+248.0625+770.0625+189.0625}{15}=$$

$$=\frac{3687}{15}=245.8$$

Тепер обчислимо залишкову оцінку дисперсії:

$$\left(\sigma^2\right)^{**}=\frac{1}{n-m}\|\overrightarrow{\eta}-F\overrightarrow{\beta^*}\|^2$$

У нашому випадку $n=16, m=2$. Перш ніж обчислювати значення статистики, знайдемо значення квадрата норми.

$$\|\overrightarrow{\eta}_{\text{зн}}-F\overrightarrow{\beta^*_{\text{зн}}}\|^2\approx 502.96075$$

Нарешті, можемо знайти значення залишкової оцінки дисперсії:

$$\left(\sigma^2\right)^{**}_{\text{зн}}=\frac{502.96075}{16-2}\approx 35.92577$$

Значення статистики F-критерію:

$$\zeta_{\text{зн}}=\frac{(D^{**}\eta)_{\text{зн}}}{(\sigma^2)^{**}_{\text{зн}}}=\frac{245.8}{35,92577}\approx 6.84$$

За таблицею Фішера знайдемо значення $t_{\text{кр}}$:

$$t_{\text{кр}}=F(n-1,n-m)=F(15,14)=2.46$$

Отже, $\zeta_{\text{зн}}>t_{\text{кр}}$ і модель можна вважати адекватною на рівні значущості 0.05.

4.Для самого малого значення параметра побудованої моделі на рівні значущості $\alpha = 0.05$ перевірити гіпотезу про його значущість.

Якщо значення якоїсь з оцінок β_j^* виявилося близьким до нуля, це може означати, що відповідний фактор не робить внеску до моделі. У нашому випадку найближчим до нуля значенням серед оцінок параметрів є $(\beta_1^*)_{\text{зн}} = 1.56844$. Саме його значущість ми і будемо перевіряти.
Висунемо основну та альтернативну гіпотези:

$$H_0: \ \beta_1 = 0$$

$$H_1: \ \beta_1 > 0$$

Критична область прицьому буде правосторонньою.
Статистика, за допомогою якої перевірятимемо гіпотезу має вигляд:

$$\gamma = \frac{\beta_j}{\sqrt{\left(\sigma^2\right)^{**} \cdot a_{jj}}} \ \sim \ St_{n-m}$$

Тут $\left(\sigma^2\right)^{**}$ - залишкова оцінка дисперсії, β_j - параметр, який перевіряємо на значущість, a_{jj} - відповідний елемент дисперсійної матриці Фішера. В нашому випадку $(\beta_1^*)_{\text{зн}} = 1.56844$, $\left(\sigma^2\right)^{**}_{\text{зн}} = 35.92577$, $a_{11} = 0.00074$. Обчислюємо значення статистики:

$$\gamma_{\text{зн}} = \frac{1.56844}{\sqrt{35.92577 \cdot 0.00074}} \approx \frac{1.56844}{0.16305} \approx 9.619$$

Значення $t_{\text{кр}}$ веремо з таблиці Стьюдента. Кількість ступенів вільності при цьому дорівнює $n - m = 16 - 2 = 14$.

$$t_{\text{кр}} = t_{0.05,14} = 1.761$$

Бачимо, що $\gamma_{\text{зн}} = 9.619 > t_{\text{кр}} = 1.761$, отже гіпотеза H_0 відхиляється на користь гіпотези H_1 на рівні значущості $\alpha = 0.05$. Таким чином, отримали, що дослідні дані не суперечать значущості параметра β_2 на рівні значущості $\alpha = 0.05$, отже модель не змінюємо.

5.Побудувати прогнозований довірчий інтервал з довірчою ймовірністю $\gamma = 0.95$ для середнього значення відклику та самого значення відклику в точці $x = 16$.

Для побудови довірчих інтервалів використаємо статистику:

$$\theta = \frac{f^*(\overrightarrow{x}) - f(\overrightarrow{x})}{\sqrt{(\sigma^2)^{**}\left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x}}} \ \sim \ St_{n-m}$$

Довірчий інтервал для середнього значення відклику має вигляд:

$$f(\overrightarrow{x}) \in \left(f^*(\overrightarrow{x}) - t\sqrt{(\sigma^2)^{**}\left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x}}\ ;\ f^*(\overrightarrow{x}) + t\sqrt{(\sigma^2)^{**}\left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x}}\right)$$

Значення t знаходимо з умови:

$$F\left\{|\theta| < t\right\} = \gamma$$

Для нашого випадку:

$$\overrightarrow{x} = (1, 16)^T$$

$$f^*(16) = -10.86056 + 25.09504 = 14.23448$$

$$\overrightarrow{(x)}^T A^{-1}\overrightarrow{x} = 0.07034$$

$$\left(\sigma^2\right)^{**}_{\text{зн}}\overrightarrow{(x)}^T A^{-1}\overrightarrow{x} = 35.92577 \cdot 0.07034 \approx 2.52702$$

$$t_{\text{кр}} = St_{\frac{1-0.95}{2}, 14} = 2.145$$

Отже, за наведеною формулою, отримуємо довірчий інтервал для середнього значення відклику в точці $x = 16$:

$$f\left(\overrightarrow{x}\right) \in (10.82466\ ;\ 17.6443) \approx (10.8\ ;\ 17.6)$$

Довірчий інтервал для значення відклику має вигляд:

$$\eta \in \left(f^*(\overrightarrow{x}) - t\sqrt{(\sigma^2)^{**}\left(1 + \left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x}\right)}\ ;\ f^*(\overrightarrow{x}) + t\sqrt{(\sigma^2)^{**}\left(1 + \left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x}\right)}\right)$$

$$1 + \left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x} = 1.07034$$

Обчислимо невідомі значення:

$$\left(\sigma^2\right)^{**}\left(1 + \left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x}\right) = 35.92577 \cdot 1.07034 \approx 38.45279$$

Отже, за наведеною формулою маємо довірчий інтервал для значення відклику в точці $x = 16$:

$$\eta_{\text{зн}} \in (0.93327\ ;\ 27.53569) \approx (0.9\ ;\ 27.5)$$

6. Висновок

Отже, під час виконання завдання було побудовано наступну однофакторну лінійну регресійну модель:

$$f(x) = \beta_0 + \beta_1 x$$

Оцінивши за допомогою метода найменших квадратів параметри, отримали наступне:

$$f^*(x) = -10.86056 + 1.56844x$$

Далі, за допомогою F - критерія було визначено, що модель є адекватною на рівні значущості $\alpha = 0.05$, це означає, що дисперсія побудованої моделі значно мешна за дисперсію моделі-константи.
Наступним кроком було перевірено значущість параметра β_1, оскільки його значення, отримане методом найменших квадратів було найближчим до нуля. Дослідні дані не суперечили значущості параметра на рівні значущості $\alpha = 0.05$, тож було вирішено не прибирати його з моделі.
Зрештою, були побудовані довірчі інтервали для середнього значення відклику та самого значення відклику в точці $x = 16$. Варто зазначити, що обсяг вибірки ($n = 16$) доволі малий, що погіршує точність досліджень. Зокрема, довірчі інтервали вийшли досить широкими.

Завдання 2:
Дана таблиця експериментальних даних.
Треба:
1.За методом найменших квадратів знайти оцінки параметрів двофакторної регресійної моделі.
2.На рівні значущості $\alpha = 0.05$ перевірити адекватність побудованої моделі.
3.Для самого малого значення параметра побудованої моделі на рівні значущості $\alpha = 0.05$ перевірити гіпотезу про його значущість.
4.Побудувати прогнозований довірчий інтервал з довірчою ймовірністю $\gamma = 0.95$ для середнього значення відклику та самого значення відклику в деякій точці (точку вибирайте самі).
5. Написати висновки

Завдання відповідно до варіанту 7:

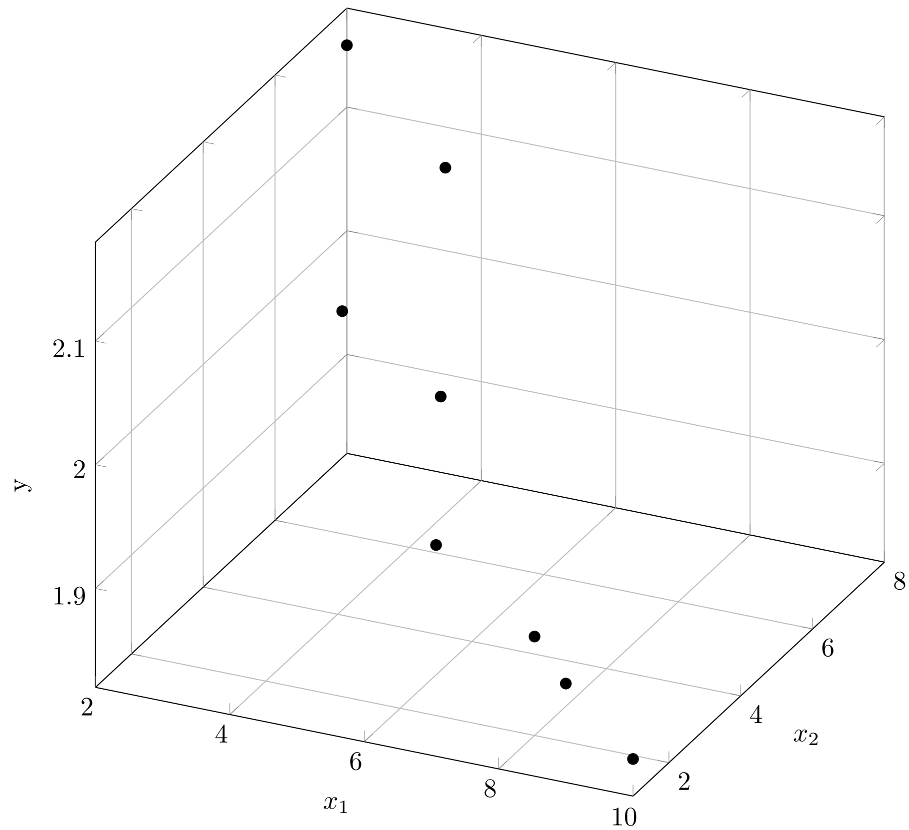
№	X_{1i}	X_{2i}	Y_i
1	10	1	1.85
2	4	7	2.15
3	4	7	2.10
4	4	7	2.05
5	9	1	1.90
6	5	5	2.00
7	5	5	1.85
8	5	5	2.10
9	8	2	1.80
10	8	2	2.00
11	3	6	2.15
12	3	6	1.85
13	2	8	2.15
14	6	3	1.90
15	6	3	1.95

1.За методом найменших квадратів знайти оцінки параметрів двофакторної регресійної моделі.

Перш ніж переходити до побудови регресійної моделі, проаналізуємо дані, які нам дані за умовою.
Бачимо, що вибірка містить значення факторів, які повторюються. Для подальшої роботи доцільно "склеїти" однакові значення. Для цього замінимо значення відклику на середнє арифметичне значень відклику "склеєних точок". Отримаємо наступне:

№	X_{1i}	X_{2i}	Y_i
1	10	1	1.85
2	4	7	2.10
3	9	1	1.90
4	5	5	1.98
5	8	2	1.90
6	3	6	2.00
7	2	8	2.15
8	6	3	1.925

Зрозуміло, що при цьому зменшився і обсяг вибірки (тепер $n = 8$). Побудуємо графічне зображення вибірки:



Розглянемо модель $y = f(\overrightarrow{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.
Запишемо матрицю плану:

$$F = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 10 & 4 & 9 & 5 & 8 & 3 & 2 & 6 \\ 1 & 7 & 1 & 5 & 2 & 6 & 8 & 3 \end{pmatrix}^T$$

Обчислимо інформаційну матрицю:

$$A = F^T F = \begin{pmatrix} 8 & 47 & 33 \\ 47 & 335 & 140 \\ 33 & 140 & 189 \end{pmatrix}$$

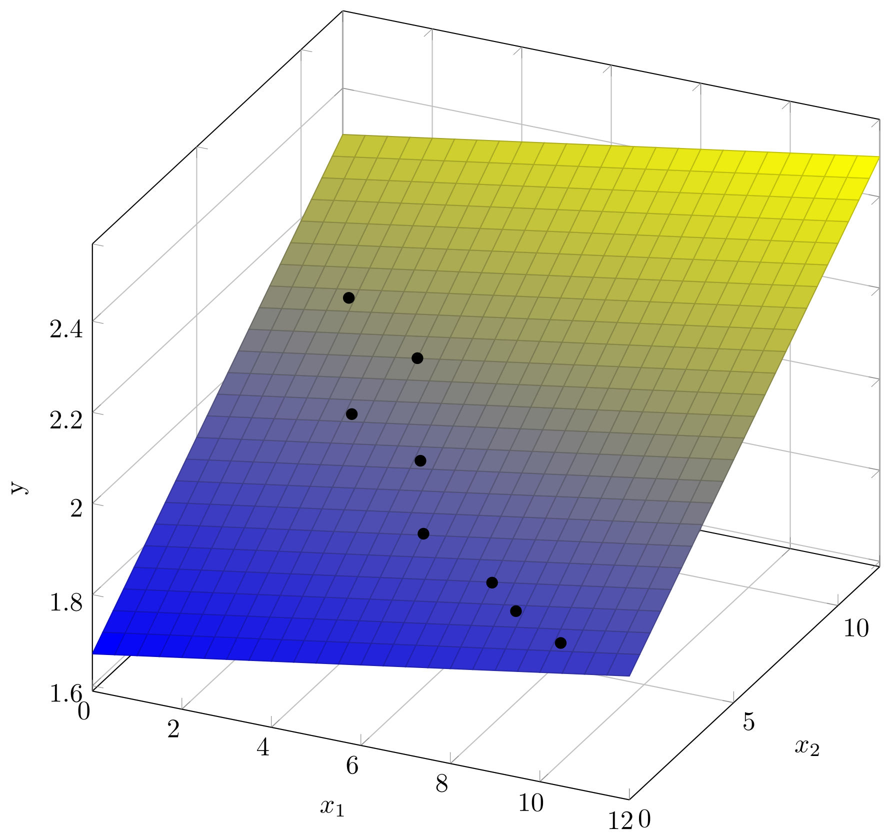
Дисперсійна матриця Фішера:

$$A^{-1} \approx \begin{pmatrix} 25.95903 & -2.53147 & -2.65736 \\ -2.53147 & 0.25119 & 0.25594 \\ -2.65736 & 0.25594 & 0.27969 \end{pmatrix}$$

$$\overrightarrow{\eta}_{\text{зн}} = \begin{pmatrix} 1.85 & 2.10 & 1.90 & 1.98 & 1.90 & 2.00 & 2.15 & 1.925 \end{pmatrix}^T$$

$$\overrightarrow{\beta^*}_{\text{зн}} = A^{-1}F^T\overrightarrow{\eta}_{\text{зн}} \approx \begin{pmatrix} 1.66998 \\ 0.01582 \\ 0.05236 \end{pmatrix}$$

Отже, маємо модель: $y = f^*(\overrightarrow{x}) = 1.66998 + 0.01582x_1 + 0.05236x_2$. Зобразимо її на рисунку:



2.На рівні значущості $\alpha = 0.05$ перевірити адекватність побудованої моделі.

Для перевірки моделі на адекватність знову використаємо F - критерій. Нагадаємо, що обсяг вибірки $n = 8$, а кількість параметрів моделі $m = 3$. Спершу обчислимо значення вибіркового середнього:

$$\overline{y} = \frac{1}{8}\left(1.85 + 2.10 + 1.90 + 1.98 + 1.90 + 2.00 + 2.15 + 1.925\right) \approx 1.98$$

Тоді значення виправленої вибіркової дисперсії:

$$(D^{**}\eta)_{\text{зн}} = \frac{1}{7}\sum_{k=1}^{8}((y_k - \overline{y}))^2 =$$

$$= \frac{(1.85-1.98)^2 + (2.10-1.98)^2 + 2\cdot(1.90-1.98)^2 + (1.98-1.98)^2 + (2-1.98)^2}{7} +$$

$$+ \frac{(2.15-1.98)^2 + (1.925-1.98)^2}{7} =$$

$$= \frac{0.0169 + 0.0144 + 0.0128 + 0 + 0.0004 + 0.0289 + 0.003025}{7} = \frac{0.076425}{7} \approx 0.01092$$

Обчислимо значення квадрата норми:

$$\|\overrightarrow{\eta}_{\text{зн}} - F\overrightarrow{\beta^*_{\text{зн}}}\|^2 \approx 0.00501$$

Нарешті, можемо знайти значення залишкової оцінки дисперсії:

$$\left(\sigma^2\right)^{**}_{\text{зн}} = \frac{0.00501}{8-3} \approx 0.001$$

Значення статистики F-критерію:

$$\zeta_{\text{зн}} = \frac{(D^{**}\eta)_{\text{зн}}}{(\sigma^2)^{**}_{\text{зн}}} = \frac{0.01092}{0.001} = 10.92$$

За таблицею Фішера знайдемо значення $t_{\text{кр}}$:

$$t_{\text{кр}} = F(n-1, n-m) = F(7,5) = 4.88$$

Отже, $\zeta_{\text{зн}} > t_{\text{кр}}$ і на рівні значущості 0.05 дослідні дані не суперечать адекватності моделі.

3.Для самого малого значення параметра побудованої моделі на рівні значущості $\alpha = 0.05$ перевірити гіпотезу про його значущість.

Перевірятимемо гіпотезу про значущість параметра β_1.
Висунемо основну та альтернативну гіпотези:

$$H_0 : \ \beta_1 = 0$$

$$H_1 : \ \beta_1 > 0$$

Критична область при цьому буде правосторонньою.
Статистика, за допомогою якої перевірятимемо гіпотезу має вигляд:

$$\gamma = \frac{\beta_j}{\sqrt{\left(\sigma^2\right)^{**} \cdot a_{jj}}} \ \sim \ St_{n-m}$$

Тут $\left(\sigma^2\right)^{**}$ - залишкова оцінка дисперсії, β_j - параметр, який перевіряємо на значущість, a_{jj} - відповідний елемент дисперсійної матриці Фішера. В нашому випадку $(\beta_1^*)_{\text{зн}} = 0.01582$, $\left(\sigma^2\right)^{**}_{\text{зн}} = 0.001$, $a_{11} = 0.25119$. Обчислюємо значення статистики:

$$\gamma_{\text{зн}} = \frac{0.01582}{\sqrt{0.001 \cdot 0.25119}} \approx \frac{0.01582}{0.0158} \approx 1$$

Значення $t_{\text{кр}}$ веремо з таблиці Стьюдента. Кількість ступенів вільності при цьому дорівнює $n - m = 8 - 3 = 5$.

$$t_{\text{кр}} = t_{0.05,5} = 2.015$$

Бачимо, що $\gamma_{\text{зн}} = 1 < t_{\text{кр}} = 2.015$, це означає, що дослідні дані не суперечать гіпотезі H_0 на рівні значущості $\alpha = 0.05$. У нашому випадку це означає, що параметр β_2 не є значущим, отже з регресійної моделі відповідний фактор можна прибрати.
Нова модель матиме вигляд: $f^{\sim}(\overrightarrow{x}) = \beta_0^{\sim} + \beta_1^{\sim} x_2$. Знайдемо значення параметрів:
Матриця плану:

$$F = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 7 & 1 & 5 & 2 & 6 & 8 & 3 \end{pmatrix}^T$$

Інформаційна матриця:

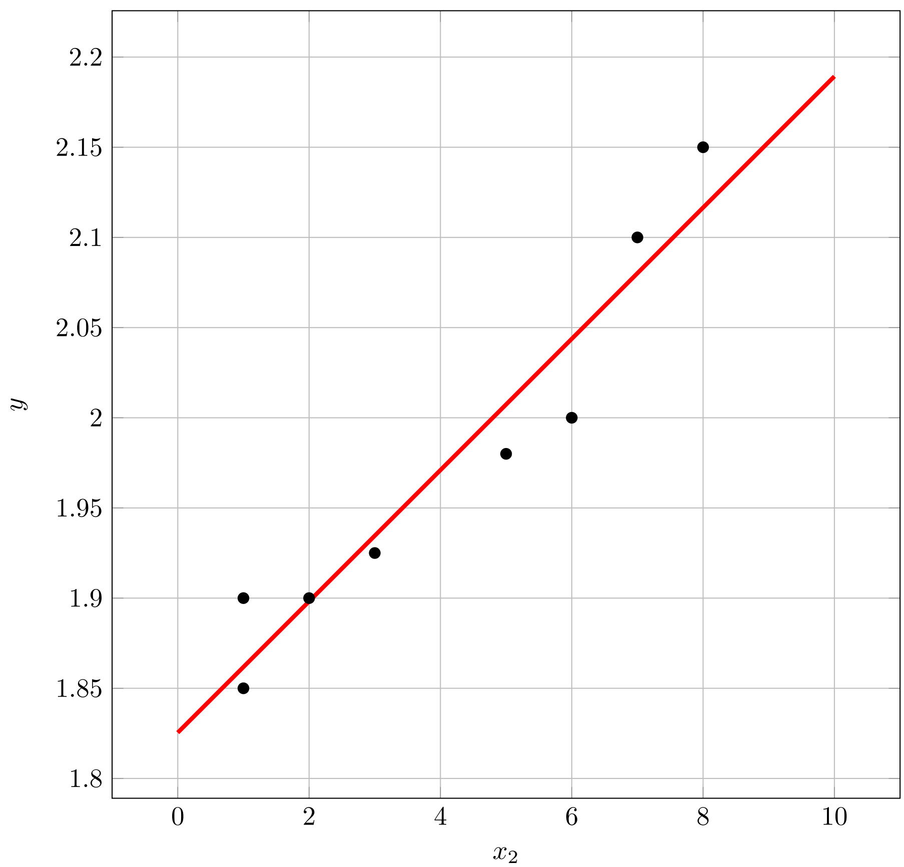
$$A = F^T F = \begin{pmatrix} 8 & 33 \\ 33 & 189 \end{pmatrix}$$

Дисперсійна матриця Фішера:

$$A^{-1} = \begin{pmatrix} 0.44681 & -0.07801 \\ -0.07801 & 0.01891 \end{pmatrix}$$

$$\overrightarrow{\beta^*}_{\text{зн}} = A^{-1}F^T\overrightarrow{\eta}_{\text{зн}} \approx \begin{pmatrix}1.82541\\0.03639\end{pmatrix}$$

Отже, маємо модель: $y = f^{*\sim}(\overrightarrow{x}) = 1.82541 + 0.03639x_2$. Зобразимо її на рисунку (оскільки модель фактично залежить лише від фактора x_2 побудовано буде проекцію на площину OX_2Y):



Тепер перевіримо модель на адекватність на рівні значущості $\alpha = 0.05$. Зауважимо, що обсяг вибірки не змінився, а кількість параметрів зменшилась на 1, тобто тепер $m = 2$.

$$(D^{**}\eta)_{\text{зн}} \approx 0.01092$$

Обчислимо значення квадрата норми:

$$\|\overrightarrow{\eta}_{\text{зн}} - F\overrightarrow{\beta^*_{\text{зн}}}\|^2 \approx 0.00587$$

Можемо знайти значення залишкової оцінки дисперсії:

$$\left(\sigma^2\right)^{**}_{\text{зн}} = \frac{0.00501}{8-2} \approx 0.00098$$

Значення статистики F-критерію:

$$\zeta_{\text{зн}} = \frac{(D^{**}\eta)_{\text{зн}}}{(\sigma^2)^{**}_{\text{зн}}} = \frac{0.01092}{0.00098} \approx 11.14$$

За таблицею Фішера знайдемо значення $t_{\text{кр}}$:

$$t_{\text{кр}} = F(n-1, n-m) = F(7,6) = 4.21$$

Отже, $\zeta_{\text{зн}} > t_{\text{кр}}$ і модель можна вважати адекватною на рівні значущості 0.05.
Тепер перевіримо значущість параметра $\beta_1^{\sim}$.
Висунемо основну та альтернативну гіпотези:

$$H_0: \ \beta_1^{\sim} = 0$$

$$H_1: \ \beta_1^{\sim} > 0$$

Критична область при цьому буде правосторонньою.
Статистика, за допомогою якої перевірятимемо гіпотезу має вигляд:

$$\gamma = \frac{\beta_j}{\sqrt{\left(\sigma^2\right)^{**} \cdot a_{jj}}} \ \sim \ St_{n-m}$$

Тут $\left(\sigma^2\right)^{**}$ - залишкова оцінка дисперсії, β_j - параметр, який перевіряємо на значущість, a_{jj} - відповідний елемент дисперсійної матриці Фішера. В нашому випадку $\left(\beta_1^{*\sim}\right)_{\text{зн}} = 0.03639$, $\left(\sigma^2\right)^{**}_{\text{зн}} = 0.00098$, $a_{11} = 0.01891$. Обчислюємо значення статистики:

$$\gamma_{\text{зн}} = \frac{0.03639}{\sqrt{0.00098 \cdot 0.01891}} \approx \frac{0.03639}{0.0043} \approx 8.463$$

Значення $t_{\text{кр}}$ веремо з таблиці Стьюдента. Кількість ступенів вільності при цьому дорівнює $n - m = 8 - 2 = 6$.

$$t_{\text{кр}} = t_{0.05,6} = 1.943$$

Бачимо, що $\gamma_{\text{зн}} = 8.463 > t_{\text{кр}} = 1.943$, отже гіпотеза H_0 відхиляється на користь гіпотези H_1 на рівні знаущості $\alpha = 0.05$. Таким чином, отримали, що параметр $\beta_1^{\sim}$ є значущим, отже модель не змінюємо.

4.Побудувати прогнозований довірчий інтервал з довірчою ймовірністю $\gamma = 0.95$ для середнього значення відклику та самого значення відклику в деякій точці (точку вибирайте самі).

Інтервали будуватимемо для точки $\overrightarrow{x_0} = (10, 1)^T$.
Довірчий інтервал для середнього значення відклику має вигляд:

$$f(\overrightarrow{x}) \in \left(f^*(\overrightarrow{x}) - t\sqrt{(\sigma^2)^{**}\left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x}}\ ;\ f^*(\overrightarrow{x}) + t\sqrt{(\sigma^2)^{**}\left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x}} \right)$$

Значення t знаходимо з умови:

$$F\left\{|\theta| < t\right\} = \gamma$$

Для нашого випадку:

$$\overrightarrow{x} = (1, 1)^T$$

$$f^{*^{\sim}}(1, 1) = 1.82541 + 0.03639 = 1.8618$$

$$\overrightarrow{(x)^T} A^{-1}\overrightarrow{x} = 0.3097$$

$$\left(\sigma^2\right)^{**}_{\text{зн}} \overrightarrow{(x)^T} A^{-1}\overrightarrow{x} = 0.00098 \cdot 0.3097 \approx 0.0003$$

$$t_{\text{кр}} = St_{\frac{1-0.95}{2}, 6} = 2.447$$

Отже, за наведеною формулою, отримуємо довірчий інтервал для середнього значення відклику в точці $\overrightarrow{x_0} = (10, 1)^T$:

$$f^{\sim}\left(\overrightarrow{x}\right) \in (1.81917\ ;\ 1.90443) \approx (1.8\ ;\ 1.9)$$

Довірчий інтервал для значення відклику має вигляд:

$$\eta \in \left(f^*(\overrightarrow{x}) - t\sqrt{(\sigma^2)^{**}\left(1 + \left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x}\right)}\ ;\ f^*(\overrightarrow{x}) + t\sqrt{(\sigma^2)^{**}\left(1 + \left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x}\right)} \right)$$

$$1 + \left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x} = 1.3097$$

Обчислимо невідомі значення:

$$\left(\sigma^2\right)^{**}\left(1 + \left(\overrightarrow{x}\right)^T A^{-1}\overrightarrow{x}\right) = 0.00098 \cdot 1.3097 \approx 0.00128$$

Отже, за наведеною формулою маємо довірчий інтервал для значення відклику в точці $\overrightarrow{x_0} = (10, 1)^T$:

$$\eta_{\text{зн}} \in (1.77413\ ;\ 1.94947) \approx (1.77\ ;\ 1.95)$$

6. Висновок

Під час виконання завдання було побудовано наступну модель:

$$f(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Оцінивши за допомогою МНК параметри моделі отримали:

$$f^*(\overrightarrow{x}) = 1.66998 + 0.01582x_1 + 0.05236x_2$$

Перевіривши цю модель на адекватність за допомогою F - критерію, отримали, що дослідні дані не суперечать адекватності моделі на рівні значущості $\alpha = 0.05$.
Однак, під час перевірки значущості найменшого параметра, виявили, що при тому самому значенні α, значущість параметра β_1 не підтверджується. Було зроблено висновок, що даний параметр можна прибрати з моделі. Тож отримали нову модель:

$$f^{\sim}(\overrightarrow{x}) = \beta_0^{\sim} + \beta_1^{\sim} x_2$$

Зрозуміло, що при зміні моделі необхідно знові провести її дослідження. Отже, спершу за МНК було знайдено значення параметрів:

$$f^{*^{\sim}}(\overrightarrow{x}) = 1.82541 + 0.03639x_2$$

Далі, було перевірено адекватність моделі на рівні значущості $\alpha = 0.05$. Виявилося, що на заданому рівні значущості модель можна вважати адекватною, тож наступним кроком було перевірено значущість найменшого параметра $\beta_1^{\sim}$ на тому самому рівні значущості. Після проведення всіх необхідних обчислень, було виявлено, що на рівні значущості $\alpha = 0.05$ дослідні дані не суперечать значущості параметра, отже надалі робота велася з новою моделлю.
Зрештою, знову було побуовано довірчі інтервали для середнього значення відклику та самого значення відклику у точці $\overrightarrow{x_0} = (10, 1)^T$. Інтервали вийшли доволі точними на відміну від однофакторної моделі.