

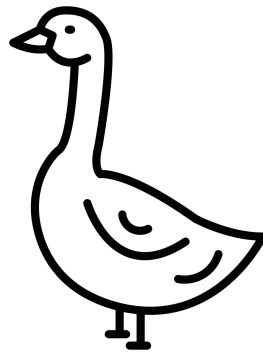
Розрахункова робота № 2

Регресійний аналіз

студента КА-02

Романовича Володимира

Варіант 9



Київ 2022

Завдання розрахункової роботи

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 найменшого 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 деякій точці.
6. Написати висновки

Задача 1

1. Провести аналіз вибірки та вибрати підходящу лінійну регресійну модель

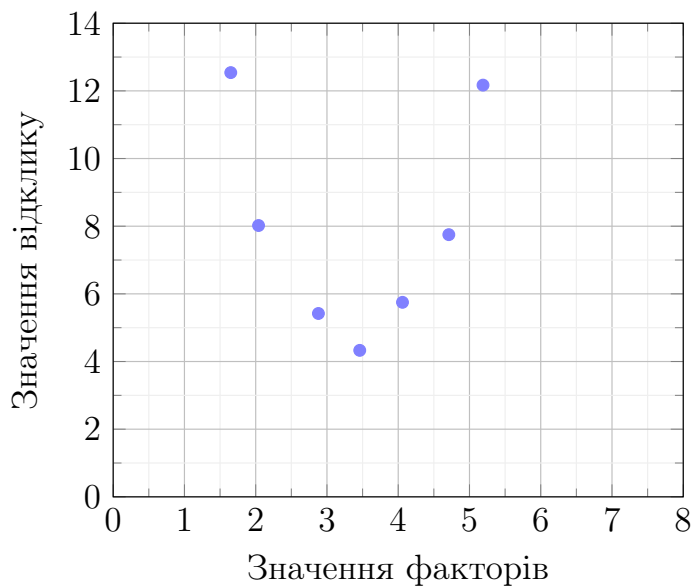
У результаті проведення деякого експерименту 7 разів отримали таблицю значень:

X	1.65	2.04	2.88	3.46	4.06	4.71	5.19
Y	12.54	8.02	5.42	4.33	5.75	7.75	12.17

Маємо вектор значень фактору $\vec{x} = (1.65, 2.04, 2.88, 3.46, 4.06, 4.71, 5.19)^\top$

та вектор значень відкликів: $\vec{\eta}_{\text{зн}} = (12.54, 8.02, 5.42, 4.33, 5.75, 7.75, 12.17)^\top$

Зобразимо результат графічно:



Можемо замітити що графік залежності значень відкликів від значень фактору нагадує параболу. Припустимо що $f(x) = \mathbb{E}(\eta/\xi = x) = \beta_0 + \beta_1 x + \beta_2 x^2$ – поліноміальна регресія.

2. За методом найменших квадратів знайти оцінки параметрів вибраної моделі

Позначимо $\varepsilon_i = \eta_i - f(x_i)$, $\vec{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top$.

Тоді вектор відкликів можна записати як $\vec{\eta} = F\vec{\beta} + \vec{\varepsilon}$, де F – матриця плану.

Припустимо, що $\vec{\varepsilon} \sim N(\vec{0}, \sigma^2 \mathbb{I})$. Відомо що в такому випадку методом найменших квадратів ми отримуємо наступну оцінку параметрів $\vec{\beta}$: $\vec{\beta}^* = A^{-1} F^\top \vec{\eta}$, де $A^{-1} = (F^\top F)^{-1}$ – дисперсійна матриця.

При чому на лекції було доведено незміщеність та конзистентність даної оцінки, а за теоремою Гаусса-Маркова й ефективність

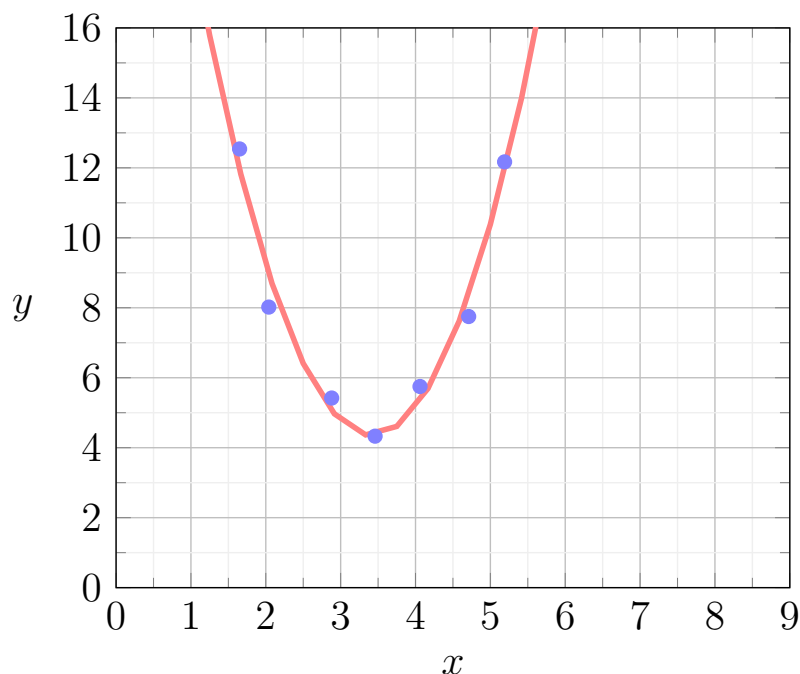
Знайдемо $\vec{\beta}_{\text{зн}}^*$:

$$F = \begin{pmatrix} 1 & 1.65 & 2.7225 \\ 1 & 2.04 & 4.1616 \\ 1 & 2.88 & 8.2944 \\ 1 & 3.46 & 11.9716 \\ 1 & 4.06 & 16.4836 \\ 1 & 4.71 & 22.1841 \\ 1 & 5.19 & 26.9361 \end{pmatrix}, A = F^T F \approx \begin{pmatrix} 7 & 23.99 & 92.75 \\ 23.99 & 92.75 & 389.5 \\ 92.75 & 389.5 & 1726.24 \end{pmatrix}, A^{-1} \approx \begin{pmatrix} 10.96 & -6.89 & 0.97 \\ -6.89 & 4.54 & -0.65 \\ 0.97 & -0.65 & 0.10 \end{pmatrix}$$

Отже

$$\vec{\beta}_{\text{зн}}^* = A^{-1} F^T \vec{\eta}_{\text{зн}} \approx \begin{pmatrix} 10.96 & -6.89 & 0.97 \\ -6.89 & 4.54 & -0.65 \\ 0.97 & -0.65 & 0.10 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1.65 & 2.7225 \\ 1 & 2.04 & 4.1616 \\ 1 & 2.88 & 8.2944 \\ 1 & 3.46 & 11.9716 \\ 1 & 4.06 & 16.4836 \\ 1 & 4.71 & 22.1841 \\ 1 & 5.19 & 26.9361 \end{pmatrix}^T \cdot \begin{pmatrix} 12.54 \\ 8.02 \\ 5.42 \\ 4.33 \\ 5.75 \\ 7.75 \\ 12.17 \end{pmatrix} \approx \begin{pmatrix} 32.7375 \\ -16.5924 \\ 2.4244 \end{pmatrix}$$

Отримали таку модель: $y = f_{\text{зн}}^*(x) = 32.7375 - 16.5924x + 2.4244x^2$. Знизу наведено її графік разом з діаграмою розмаху.



Перевіримо дану модель на адекватність.

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

Розглянемо вектор $\vec{\varepsilon}$, кожна його координата є нормально розподіленою з дисперсією σ^2 .

Ми знаємо, що $(\sigma^2)^* = \frac{1}{n-m} \|\vec{\eta} - F\vec{\beta}^*\|^2$ – незміщена оцінка дисперсії ε_i .

Також відомо, що $\frac{n-m}{\sigma^2}(\sigma^2)^* = \frac{1}{\sigma^2}\|\vec{\eta} - F\vec{\beta}^*\|^2 \sim \chi_{n-m}^2$.

Оскільки $\eta_i = \varepsilon_i + f(x_i)$ розподілені нормально з дисперсією σ^2 , то $\frac{n-1}{\sigma^2}D^{**}\eta \sim \chi_{n-1}^2$

Звідси маємо, що статистика ζ має розподіл Фішера-Снедекора.

$$\zeta = \frac{\frac{n-1}{\sigma^2}D^{**}\eta \cdot \frac{1}{n-1}}{\frac{n-m}{\sigma^2}(\sigma^2)^* \cdot \frac{1}{n-m}} = \frac{\frac{1}{n-1} \sum_{k=1}^n (\eta_k - \bar{\eta})^2}{\frac{1}{n-m} \|\vec{\eta} - F\vec{\beta}^*\|^2} \sim F(n-1, n-m)$$

Розглянемо тепер модель $f_c(x) = \beta_c$. На лекції було доведено, що $\beta_c^* = \bar{\eta}$

Незміщена оцінка дисперсії похибок спостережень в такому випадку

$$(\sigma_c^2)^* = \frac{1}{n-1} \|\vec{\eta} - \bar{\eta} \cdot F\|^2 = \frac{1}{n-1} \sum_{k=1}^n (\eta_k - \bar{\eta})^2 = D^{**}\eta$$

Висунемо гіпотезу $H_0 : \sigma_c^2 = \sigma^2$ та альтернативну гіпотезу $H_1 : \sigma_c^2 > \sigma^2$.

Оскільки $(\sigma_c^2)^* = D^{**}\eta$, для перевірки гіпотези використовуватимемо статистику ζ .

Маємо $n = 7$, $m = 3$, знайдемо $(D^{**}\eta)_{\text{зн}}$, $(\sigma^2)_{\text{зн}}^*$:

$$(\bar{\eta})_{\text{зн}} = \frac{1}{7} \sum_{k=1}^7 \eta_k \approx 7.997, \quad (D^{**}\eta)_{\text{зн}} = \frac{1}{6} \sum_{k=1}^7 (\eta_k - 7.997)^2 \approx 10.5419$$

$$\vec{\varepsilon}_{\text{зн}} = \vec{\eta}_{\text{зн}} - F\vec{\beta}_{\text{зн}}^* \approx \begin{pmatrix} 12.54 \\ 8.02 \\ 5.42 \\ 4.33 \\ 5.75 \\ 7.75 \\ 12.17 \end{pmatrix} - \begin{pmatrix} 1 & 1.65 & 2.7225 \\ 1 & 2.04 & 4.1616 \\ 1 & 2.88 & 8.2944 \\ 1 & 3.46 & 11.9716 \\ 1 & 4.06 & 16.4836 \\ 1 & 4.71 & 22.1841 \\ 1 & 5.19 & 26.9361 \end{pmatrix} \cdot \begin{pmatrix} 32.7375 \\ -16.5924 \\ 2.4244 \end{pmatrix} \approx \begin{pmatrix} 0.5796 \\ -0.9582 \\ 0.3600 \\ -0.0213 \\ 0.4154 \\ -0.6196 \\ 0.2442 \end{pmatrix}$$

$$\|\vec{\varepsilon}_{\text{зн}}\|^2 = \sum_{k=1}^7 \varepsilon_k^2 \approx 2.0003, \quad (\sigma^2)_{\text{зн}}^* = \frac{1}{7-3} \|\vec{\varepsilon}_{\text{зн}}\|^2 \approx 0.5001$$

Отже

$$\zeta_{\text{зн}} = \frac{(D^{**}\eta)_{\text{зн}}}{(\sigma^2)_{\text{зн}}^*} \approx 21.08$$

$$t_{\text{кр}} = t_{0.05, 6, 4} = 6.09, \quad \zeta_{\text{зн}} > t_{\text{кр}}$$

Оскільки критична область правостороння, то на рівні значущості 0.05 гіпотеза H_0 не справджується, що за F-критерієм означає адекватність моделі $f_{\text{зн}}^*(x) = 32.7375 - 16.5924x + 2.4244x^2$ на цьому рівні значущості.

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

Нагадаємо, що $\vec{\beta}_{\text{зн}}^* \approx \begin{pmatrix} 32.7375 \\ -16.5924 \\ 2.4244 \end{pmatrix}$, бачимо що найменшим за модулем значенням є $(\beta_2^*)_{\text{зн}} \approx 2.4244$.

Перевіримо цей параметр на значущість. Нехай $H_0 : \beta_2 = 0$, $H_1 : \beta_2 > 0$, критична область правостороння.

$\beta_2^* \sim N(\beta_2, \sigma^2 \cdot a_{22})$, тому обираємо статистику:

$$\gamma = \frac{\beta_2^*}{\sqrt{(\sigma^2)^* \cdot a_{22}}} \sim St_4$$

$$\gamma_{\text{зн}} \approx \frac{2.4244}{\sqrt{0.50001 \cdot 0.1}} \approx 10.84 > t_{0.05,4} \approx 2.132$$

Отже на рівні значущості 0.05 параметр β_2 є значущим

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

Позначимо $\vec{x} = (1, x, x^2)^\top$, тоді

$$f^*(x) = \vec{x}^\top \vec{\beta}^*$$

$$\mathbb{E}f^*(x) = \mathbb{E}\vec{x}^\top \vec{\beta}^* = \vec{x}^\top \mathbb{E}\vec{\beta}^* = \vec{x}^\top \vec{\beta} = f(x)$$

$$\mathbb{D}f^*(x) = \sigma^2 \cdot \vec{x}^\top \cdot A^{-1} \cdot \vec{x}$$

$$\frac{f^*(x) - f(x)}{\sqrt{\mathbb{D}f^*(x)}} \sim N(0, 1), \quad \frac{n-m}{\sigma^2} (\sigma^2)^* \sim \chi_{n-m}^2$$

Тому

$$\zeta = \frac{\frac{f^*(x) - f(x)}{\sqrt{\mathbb{D}f^*(x)}}}{\sqrt{\frac{n-m}{\sigma^2} (\sigma^2)^* \cdot \frac{1}{n-m}}} = \frac{f^*(x) - f(x)}{\sqrt{(\sigma^2)^* \cdot \vec{x}^\top A^{-1} \vec{x}}} \sim St_{n-m}$$

Отже маємо довірчий інтервал для середнього значення відклику:

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

З аналогічних міркувань, враховуючи що $\mathbb{E}(\eta - f^*(x)) = 0$, $\mathbb{D}(\eta - f^*(x)) = \sigma^2(1 + \vec{x}^\top A^{-1} \vec{x})$, отримуємо довірчий інтервал для значення відклику:

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

Обчислимо ці довірчі інтервали в точці $x=5$, тоді $\vec{x} = (1, 5, 25)^\top$

$\zeta \sim St_4$, знайдемо таке t , для якого $P\{|\zeta| < t\} = 0.95$:

З таблиці $t = t_{0.025,4} \approx 2.776$

$$f_{\text{зн}}^*(5) = (\beta_0^*)_{\text{зн}} + (\beta_1^*)_{\text{зн}} \cdot 5 + (\beta_2^*)_{\text{зн}} \cdot 25 \approx 10.38457$$

$$(\vec{x}^\top A^{-1} \vec{x})|_{x=5} \approx (1, 5, 25) \cdot \begin{pmatrix} 10.959813138731 & -6.891022902854 & 0.9659642070936 \\ -6.891022902854 & 4.538129338921 & -0.653692863812 \\ 0.9659642070936 & -0.653692863812 & 0.0961721539816 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 5 \\ 25 \end{pmatrix} \approx 0.48541$$

$$(\sigma^2)_{\text{зн}}^* \approx 0.500075$$

$$f(5) \in (f_1, f_2), \text{ де } f_{1,2} = 10.38457 \pm 2.776 \cdot \sqrt{0.500075 \cdot 0.48541} \implies f(5) \in (9.01687; 11.75227)$$

$$\eta \in (\eta_1, \eta_2), \text{ де } \eta_{1,2} = 10.38457 \pm 2.776 \cdot \sqrt{0.500075 \cdot (1 + 0.48541)} \implies \eta \in (7.99202; 12.77712)$$

Отже з ймовірністю 0.95 при $x=5$: $\eta \in (7.99202; 12.77712)$, $f(x) \in (9.01687; 11.75227)$

6. Висновок

Побудувавши діаграму розсіювання заданої вибірки, ми побачили що розташування точок нагадує параболу, тому ми припустили що наша модель є поліноміальною регресією вигляду $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$. В результаті дослідження та перевірки даної моделі знайшли значення найкращих оцінок параметрів та прийшли до висновку що модель є адекватною, та її параметри є значущими на заданому рівні значущості. Також були побудовані прогнозовані довірчі інтервали для значення відклику, та середнього значення відклику в точці $x=5$.

Задача 2

1. Провести аналіз вибірки та вибрати підходящу лінійну регресійну модель

У результаті проведення деякого експерименту 15 разів отримали таблицю значень:

X_{1i}	8	8	3	3	4	4	4	7	6	6	6	4	3	3	5
X_{2i}	1	1	7	7	6	6	6	3	4	4	4	7	6	6	7
Y	21.5	20.5	19.5	18.5	20	21	19.8	20.5	21.5	19.5	19	18	21	20.5	20

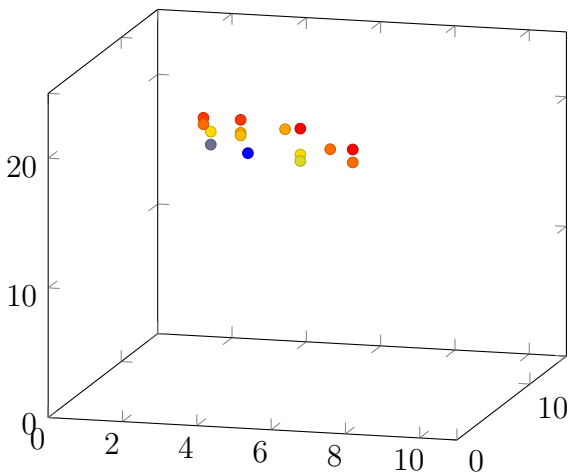
Маємо вектори значень двох факторів та вектор значень відклику:

$$\vec{x}_1 = (8, 8, 3, 3, 4, 4, 4, 7, 6, 6, 6, 6, 4, 3, 3, 5)^\top$$

$$\vec{x}_2 = (1, 1, 7, 7, 6, 6, 6, 3, 4, 4, 4, 4, 7, 6, 6, 7)^\top$$

$$\vec{\eta}_{\text{зн}} = (21.5, 20.5, 19.5, 18.5, 20, 21, 19.8, 20.5, 21.5, 19.5, 19, 18, 21, 20.5, 20)^\top$$

Зобразимо результат графічно:



Судячи з розміщення точок я вважаю доцільним припустити що функція регресії має вигляд площини: $f(\vec{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

2. За методом найменших квадратів знайти оцінки параметрів вибраної моделі

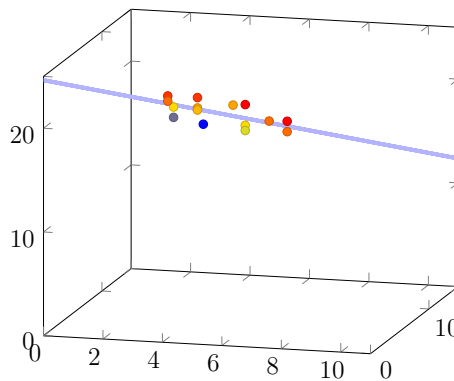
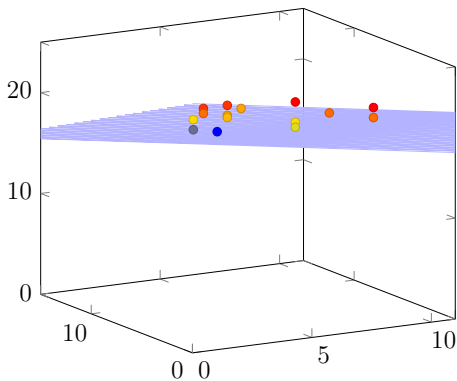
Як ми вже знаємо, за припущення що $\vec{\varepsilon} \sim N(0, \sigma^2 \mathbb{I})$, де $\vec{\varepsilon} = \vec{\eta} - F\vec{\beta}$, незміщеною, конзистентною та ефективною оцінкою параметрів є вектор $\vec{\beta}^* = A^{-1}F^\top \vec{\eta}$. Знайдемо значення цієї оцінки.

$$F = \begin{pmatrix} 1 & 8 & 1 \\ 1 & 8 & 1 \\ 1 & 3 & 7 \\ 1 & 3 & 7 \\ 1 & 4 & 6 \\ 1 & 4 & 6 \\ 1 & 4 & 6 \\ 1 & 7 & 3 \\ 1 & 6 & 4 \\ 1 & 6 & 4 \\ 1 & 6 & 4 \\ 1 & 4 & 7 \\ 1 & 3 & 6 \\ 1 & 3 & 6 \\ 1 & 5 & 7 \end{pmatrix}, \quad A = F^{\top} F = \begin{pmatrix} 15 & 74 & 75 \\ 74 & 410 & 322 \\ 75 & 322 & 435 \end{pmatrix}, \quad A^{-1} \approx \begin{pmatrix} 12.6983 & -1.3673 & -1.1772 \\ -1.3673 & 0.1531 & 0.1224 \\ -1.1772 & 0.1224 & 0.1146 \end{pmatrix}$$

$$\vec{\beta}_{\text{зн}}^* = A^{-1} F^{\top} \vec{\eta}_{\text{зн}} \approx \begin{pmatrix} 12.6983 & -1.3673 & -1.1772 \\ -1.3673 & 0.1531 & 0.1224 \\ -1.1772 & 0.1224 & 0.1146 \end{pmatrix} \cdot \begin{pmatrix} 1 & 8 & 1 \\ 1 & 8 & 1 \\ 1 & 3 & 7 \\ 1 & 3 & 7 \\ 1 & 4 & 6 \\ 1 & 4 & 6 \\ 1 & 4 & 6 \\ 1 & 7 & 3 \\ 1 & 6 & 4 \\ 1 & 6 & 4 \\ 1 & 6 & 4 \\ 1 & 4 & 7 \\ 1 & 3 & 6 \\ 1 & 3 & 6 \\ 1 & 5 & 7 \end{pmatrix}^{\top} \cdot \begin{pmatrix} 21.5 \\ 20.5 \\ 19.5 \\ 18.5 \\ 20 \\ 21 \\ 19.8 \\ 20.5 \\ 21.5 \\ 19.5 \\ 19 \\ 18 \\ 21 \\ 20.5 \\ 20 \end{pmatrix} \approx \begin{pmatrix} 24.70581632653 \\ -0.383673469388 \\ -0.551938775510 \end{pmatrix}$$

Отримали модель $y = f_{\text{зн}}^*(\vec{x}) \approx 24.70581632653 - 0.383673469388 \cdot x_1 - 0.551938775510 \cdot x_2$

Побудуємо графічне зображення цієї моделі поверх діаграми розсіювання з 2 ракурсів:



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

Аналогічно до перевірки на адекватність в першому завданні маємо статистику:

$$\zeta = \frac{\mathbb{D}^{**}\eta}{(\sigma^2)^*} = \frac{\frac{1}{n-1} \sum_{k=1}^n (\eta_k - \bar{\eta})^2}{\frac{1}{n-m} \|\vec{\eta} - F\vec{\beta}^*\|^2} \sim F(n-1, n-m)$$

Знайдемо значення цієї статистики:

$$n = 15, m = 3, \bar{\eta} = \frac{1}{15} \sum_{k=1}^{15} (\eta_k)_{\text{зн}} \approx 20.0533$$

$$(D^{**}\eta)_{\text{зн}} = \frac{1}{14} \sum_{k=1}^{15} ((\eta_k)_{\text{зн}} - 20.0533)^2 \approx 1.0712381$$

$$(\sigma^2)_{\text{зн}}^* = \frac{1}{12} \|\vec{\eta}_{\text{зн}} - F\vec{\beta}_{\text{зн}}^*\| \approx 0.8695077$$

З таблиці $t_{\text{кр}} = t_{0.05, 14, 12} \approx 2.64$

$$\zeta_{\text{зн}} = \frac{(D^{**}\eta)_{\text{зн}}}{(\sigma^2)_{\text{зн}}^*} \approx 1.232$$

Як бачимо, $\zeta_{\text{зн}} < t_{\text{кр}}$, отже наша модель не є адекватною на рівні значущості 0.05.

Такий результат був досить очікуваний, оскільки у випадку двофакторної регресії 15 значень це досить мала за обсягом вибірка, тому часто або неможливо або дуже важко побудувати адекватну модель регресії на рівні значущості 0.05 з таким обсягом вибірки.

Проте ця модель буде адекватною на рівні значущості 0.36. Якщо нас не влаштовує настільки велика ймовірність помилки першого роду, то можемо спробувати збільшити кількість факторів. Перебравши багато моделей вигляду $\sum_{k=1}^2 \sum_{i=1}^l \beta_i \varphi_i(x_k)$, де $\varphi_i(x)$ – різні функції з перелічених

$x^a, \sin(x), e^x, \ln(x), \cos(x)$, "найадекватніша" модель з отриманих: $f_2(\vec{x}) = \beta_0 + \beta_1 x_1^{1/2} + \beta_2 x_1^{1/3} + \beta_3 x_2^2 + \beta_4 x_2^6$.

$$(f_1^*(\vec{x}))_{\text{зн}} = 207.962465 - 388.267776 x_1^{1/3} + 0.486725420 x_1^{1/2} + 208.280925 x_2^2 - 0.000116604246 x_2^6$$

Ця модель за F-критерієм є адекватною на рівні значущості 0.16.

Надалі будемо розглядати обидві моделі,

$$(f_1^*(\vec{x}))_{\text{зн}} \approx 24.70581632653 - 0.383673469388 \cdot x_1 - 0.551938775510 \cdot x_2$$

$$(f_2^*(\vec{x}))_{\text{зн}} = 207.962465 - 388.267776 x_1^{1/3} + 0.486725420 x_1^{1/2} + 208.280925 x_2^2 - 0.000116604246 x_2^6$$

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

① Для моделі $f_1(\vec{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$:

$$\vec{\beta}_{\text{зн}}^* \approx \begin{pmatrix} 24.70581632653 \\ -0.383673469388 \\ -0.551938775510 \end{pmatrix}, \text{ найменше за модулем значення } -(\beta_1^*)_{\text{зн}}, \text{ отже перевіряємо } \beta_1$$

Аналогічно до перевірки значущості параметру в першому завданні, $H_0 : \beta_1 = 0$, $H_1 : \beta_1 < 0$, критична область лівостороння. Для перевірки гіпотези використовуємо статистику:

$$\gamma = \frac{\beta_1^*}{\sqrt{(\sigma^2)^* \cdot a_{11}}} \sim St_{12}$$

$$\gamma_{\text{зн}} = \frac{-0.383673469388}{\sqrt{0.8695077 \cdot 0.1531}} \approx -1.128$$

$$t_{\text{кр}} = -t_{0.05,12} \approx 1.761$$

$\gamma_{\text{зн}} > t_{\text{кр}}$, отже ми потрапили в область прийняття гіпотези.

Отримали що фактор x_1 в нашій моделі не є значущим на рівні значущості 0.05.

Заберемо його з моделі. Отримуємо $f_3(\vec{x}) = \beta_0 + \beta_1 x_2$.

$$\text{Для такої моделі } \vec{\beta}_{\text{зн}}^* = \begin{pmatrix} 21.2783 \\ -0.245 \end{pmatrix}$$

Перевіримо її знову на адекватність:

$$\zeta \sim F(14, 13), \quad \zeta_{\text{зн}} = \frac{(D^{**}\eta)_{\text{зн}}}{(\sigma^2)_{\text{зн}}^*} \approx 1.222$$

$P\{\zeta < 1.222\} \approx 0.64 \implies \alpha \approx 0.36$, отже модель $f_3(\vec{x}) = \beta_0 + \beta_1 x_2$ адекватна на тому ж рівні значущості, що і модель $f_1(\vec{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

② Для моделі $f_2(\vec{x}) = \beta_0 + \beta_1 x_1^{1/2} + \beta_2 x_1^{1/3} + \beta_3 x_2^2 + \beta_4 x_2^6$:

$$\vec{\beta}_{\text{зн}}^* \approx \begin{pmatrix} 207.962465 \\ 208.280925 \\ -388.267776 \\ 0.486725 \\ -0.00011660 \end{pmatrix}, \text{ перевіряємо значущість параметру } \beta_3$$

$H_0 : \beta_3 = 0$, $H_1 : \beta_3 < 0$, критична область лівостороння. Для перевірки гіпотези використовуємо статистику:

$$\gamma = \frac{\beta_3^*}{\sqrt{(\sigma^2)^* \cdot a_{33}}} \sim St_{10}$$

$$a_{33} \approx 3 \cdot 10^{-9}, \quad (\sigma^2)_{\text{зн}}^* \approx 0.563407, \text{ звідси}$$

$$\gamma_{\text{зн}} \approx \frac{-0.00011660}{\sqrt{0.563407 \cdot 3 \cdot 10^{-9}}} \approx -3.802$$

$$t_{\text{кр}} = -t_{0.05,10} \approx -1.812$$

Бачимо, що $\zeta_{\text{зн}} < t_{\text{кр}}$, отже параметр є значущим на рівні значущості 0.05.

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

Візьмемо точку $\vec{x}_0 = (5, 5)^\top$

Аналогічно до першого завдання:

$$f(\vec{x}) \in \left(f^*(\vec{x}) - t\sqrt{(\sigma^2)^* \cdot \vec{X}^\top A^{-1} \vec{X}}; f^*(\vec{x}) + t\sqrt{(\sigma^2)^* \cdot \vec{X}^\top A^{-1} \vec{X}} \right)$$

$$\eta \in \left(f^*(\vec{x}) - t\sqrt{(\sigma^2)^* \cdot (1 + \vec{X}^\top A^{-1} \vec{X})}; f^*(\vec{x}) + t\sqrt{(\sigma^2)^* \cdot (1 + \vec{X}^\top A^{-1} \vec{X})} \right)$$

де $\vec{X} = (1, x_2)^\top$ для $f_3(\vec{x})$, $\vec{X} = (1, x_1^{1/2}, x_1^{1/3}, x_2^2, x_2^6)^\top$ для $f_2(\vec{x})$

① Довірчі інтервали для $f_3(\vec{x}) = \beta_0 + \beta_1 x_2$:

$$\vec{X}|_{\vec{x}=(5,5)^\top} = (1, 5)^\top, \quad (\sigma^2)_{\text{зн}}^* \approx 0.8766026, \text{ з таблиці } t \approx 2.160$$

$$A^{-1} \approx \begin{pmatrix} 0.4833 & -0.0833 \\ -0.0833 & 0.0167 \end{pmatrix}, \quad \vec{X}^\top A^{-1} \vec{X} \approx (1, 5) \cdot \begin{pmatrix} 0.4833 & -0.0833 \\ -0.0833 & 0.0167 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 5 \end{pmatrix} = \frac{2}{3}$$

$$f_{\text{зн}}^*(\vec{x})|_{\vec{x}=(5,5)^\top} \approx 21.2783 - 0.245 \cdot 5 \approx 20.05333$$

Маємо:

$$f(\vec{x}) \in (f_1, f_2), \text{ де } f_{1,2} \approx 20.05333 \pm 2.160 \cdot \sqrt{0.8766026 \cdot \frac{2}{3}} \implies f(\vec{x}) \in (18.40209, 21.70457)$$

$$\eta \in (\eta_1, \eta_2), \text{ де } \eta_{1,2} \approx 20.05333 \pm 2.160 \cdot \sqrt{0.8766026 \cdot \frac{5}{3}} \implies \eta \in (17.44249, 22.66417)$$

Отримали що з ймовірністю 0.95 у точці $\vec{x} = (5, 5)^\top$: $f(\vec{x}) \in (18.40209, 21.70457)$, $\eta \in (17.44249, 22.66417)$

② Довірчі інтервали для $f_2(\vec{x}) = \beta_0 + \beta_1 x_1^{1/2} + \beta_2 x_1^{1/3} + \beta_3 x_2^2 + \beta_4 x_2^6$:

$\vec{X}|_{x=(5,5)^\top} = (1, \sqrt{5}, \sqrt[3]{5}, 25, 15625)^\top$, $(\sigma^2)_{\text{зн}}^* \approx 0.563407$, з таблиці $t \approx 2.228$

$$A^{-1} \approx \begin{pmatrix} 8295.50255 & 9290.60940 & -17287.10284 & 21.82481 & -0.00459 \\ 9290.60940 & 10467.83783 & -19450.49101 & 25.09777 & -0.00526 \\ -17287.10284 & -19450.49101 & 36153.04934 & -46.42499 & 0.00973 \\ 21.82481 & 25.09777 & -46.42499 & 0.06551 & -1.37842 \cdot 10^{-5} \\ -0.00459 & -0.00526 & 0.00973 & -1.37842 \cdot 10^{-5} & 2.96372 \cdot 10^{-9} \end{pmatrix}$$

$$\vec{X}^\top A^{-1} \vec{X} \approx 0.24158$$

$$f_{\text{зн}}^*(\vec{x})|_{\vec{x}=(5,5)^\top} \approx 207.962465 + 208.28093 \cdot 5^{1/2} - 388.26778 \cdot 5^{1/3} + 0.486725420 \cdot 5^2 - 0.00011660425 \cdot 5^6 \approx 20.0854$$

Маємо:

$$f(\vec{x}) \in (f_1, f_2), \text{ де } f_{1,2} \approx 20.0854 \pm 2.228 \cdot \sqrt{0.563407 \cdot 0.24158} \implies f(\vec{x}) \in (19.26338, 20.90742)$$

$$\eta \in (\eta_1, \eta_2), \text{ де } \eta_{1,2} \approx 20.0854 \pm 2.228 \cdot \sqrt{0.563407 \cdot 1.24158} \implies \eta \in (18.22186, 21.94894)$$

Отримали що з ймовірністю 0.95 у точці $\vec{x} = (5, 5)^\top$: $f(\vec{x}) \in (19.26338, 20.90742)$, $\eta \in (18.22186, 21.94894)$

6. Висновок

Після аналізу даної нам вибірки ми припустили що підходяща модель регресії має вигляд $f(\vec{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, та перевіривши цю модель на адекватність ми отримали що вона є адекватною тільки на рівні значущості 0.36. Згодом було перевірено чи є параметри моделі значущими. На рівні значущості 0.05 ми отримали що β_1 не є значущим параметром. Тому ми забрали цей параметр з вибірки та заново перевірили адекватність для вже нової моделі $f_2(\vec{x}) = \beta_0 + \beta_1 x_2$, прийшли до висновку що така модель є адекватною на такому ж рівні значущості. Проте оскільки рівень значущості вийшов досить великим, ми почали шукати "більш адекватну" модель. Найкращою знайденою моделлю є $f_3(\vec{x}) = \beta_0 + \beta_1 x_1^{1/2} + \beta_2 x_1^{1/3} + \beta_3 x_2^2 + \beta_4 x_2^6$, перевіривши цю модель отримали що вона є адекватною на рівні значущості 0.16 (адекватності на нижчому рівні значущості не вдалось досягнути через малий обсяг вибірки) та її найменший параметр є значущим на заданому рівні значущості. Також побудували довірчі інтервали значення відклику та середнього значення відклику для обох моделей в точці $\vec{x} = (5, 5)^\top$