

Національний технічний університет України «Київський
політехнічний інститут імені Ігоря Сікорського»

Кафедра Математичних Методів Системного Аналізу

Розрахункова робота з дисципліни "Математическая
статистика"

Перевішив

к.ф.м.н. Каніовская І. Ю.

Виконав

студент Панченко Є. С.

Київ 2021

Зміст

1. Задача 1	3
1.1. Постановка задачі	3
1.2. Аналіз вибірки та вибір лінійної регресійної моделі	3
1.3. Знаходження оцінок параметрів за методом найменших квадратів	5
1.4. Перевірка адекватності побудованої моделі	6
1.5. Перевірка гіпотези про значущість найменшого значення параметра побудованої моделі	7
1.6. Побудова прогнозованого довірчого інтервала для середнього значення відклику та самого значення відклику	8
1.7. Висновок	9
2. Задача 2	10
2.1. Постановка задачі	10
2.2. Пошук оцінок параметрів двофакторної регресійної моделі за методом найменших квадратів	10
2.3. Перевірка адекватності побудованої моделі	11
2.4. Перевірка гіпотези про значущість найменшого значення параметра побудованої моделі	12
2.5. Побудова нової однофакторної моделі	13
2.6. Перевірка адекватності нової побудованої однофакторної моделі	15
2.7. Перевірка гіпотези про значущість найменшого значення параметра нової побудованої однофакторної моделі	16
2.8. Побудова прогнозованого довірчого інтервала для середнього значення відклику та самого значення відклику нової однофакторної моделі	17
2.9. Висновок	19

1. Задача 1

1.1. Постановка задачі

x	10	15	25	35	45	55	65	75	85	95
y	2600	2100	1300	1000	820	670	580	510	490	470

Основна мета - побудувати регресійну модель та зробити її аналіз.

1.2. Аналіз вибірки та вибір лінійної регресійної моделі

Для початку побудуємо діаграму розсіювання.

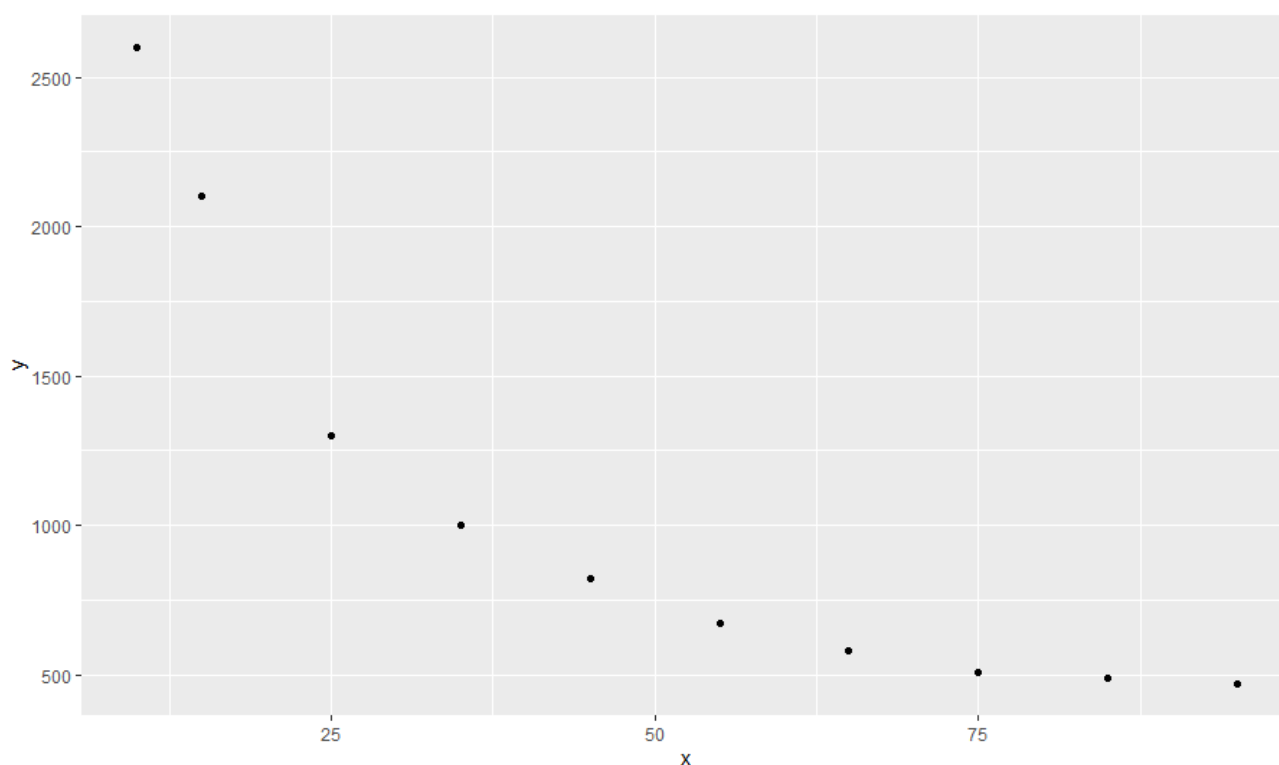


Рис. 1. Точки на площині

Помітимо, що це нагадує графік гіперболи, а тому зобразимо точки з координатами $(x, \frac{1}{y})$.

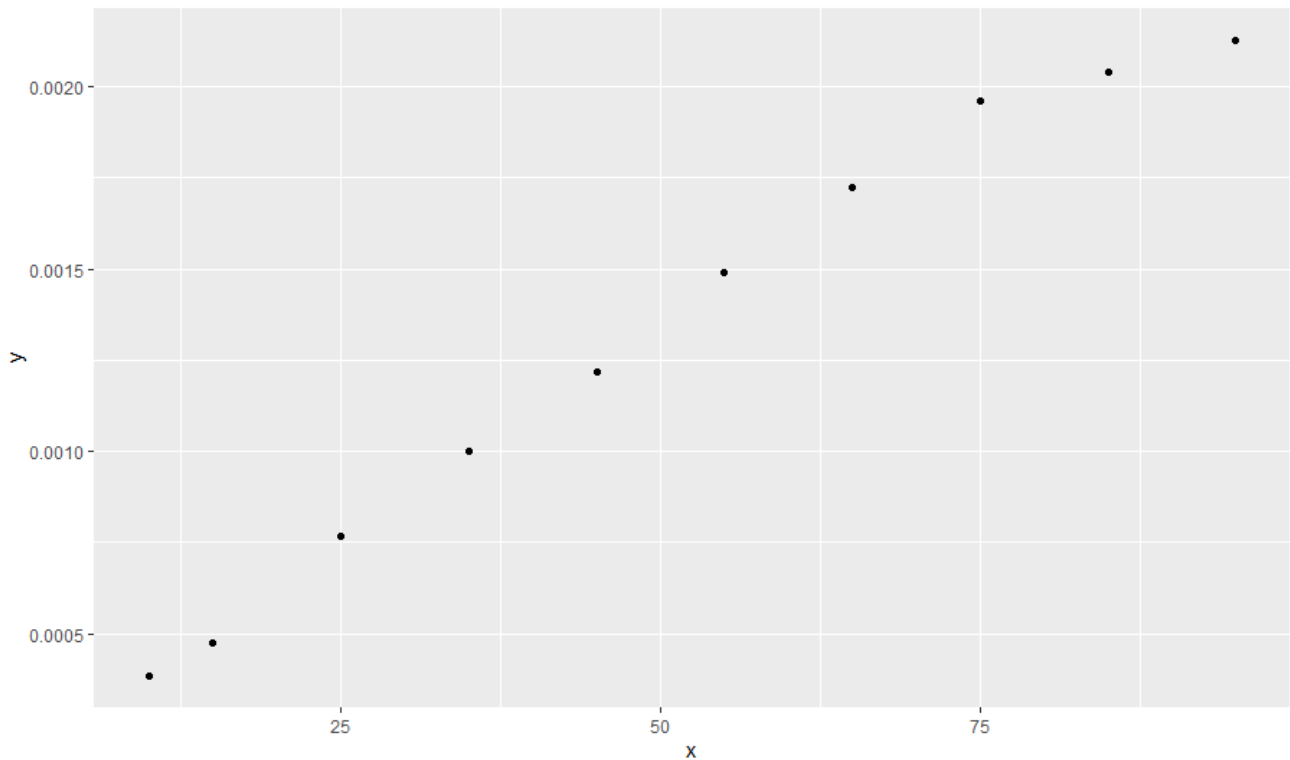


Рис. 2. Точки на площині, де координати y інверсовані

Бачимо, що точки розташовані майже на прямій. Саме тому можна обрати вигляд функції, яку оцінює регресійна модель, як

$$f(x) = \frac{1}{\beta_0 + \beta_1 x}.$$

Але цю модель можна спростити. Давайте будемо оцінювати функцію

$$g(x) = \frac{1}{f(x)}.$$

Тоді

$$g(x) = \beta_0 + \beta_1 x.$$

Одразу позначатимемо $\vec{\eta}^{(f)}$ - вектор відкликів. Також введемо позначення $\eta_i^{(g)} = \frac{1}{\eta_i^{(f)}}$. Ці позначення дозволяють нам тимчасово забути про існування функції f . Тобто можна працювати з функцією g .

1.3. Знаходження оцінок параметрів за методом найменших квадратів

Зробимо припущення, що похибка має нормальний розподіл з нульовим математичним сподіванням. Використаймо метод найменших квадратів. Він полягає у знаходженні таких значень параметрів β_0 та β_1 , щоб мінімізувати значення

$$\sum_{i=1}^{10} (\eta_i^{(g)} - g(x_i))^2.$$

Відомо, що оцінкою методом найменших квадратів параметрів лінійної регресії з припущенням, що похибка має нормальний розподіл з нульовим математичним сподіванням, є вектор

$$\vec{\beta}^* = (F^T F)^{-1} F^T \vec{\eta}^{(g)}, \quad \text{де } F - \text{матриця плану.}$$

В умовах нашої задачі

$$F = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 10 & 15 & 25 & 35 & 45 & 55 & 65 & 75 & 85 & 95 \end{pmatrix}^T$$

і

$$\vec{\eta}^{(g)} = (0.384 \ 0.476 \ 0.769 \ 1 \ 1.219 \ 1.492 \ 1.724 \ 1.96 \ 2.04 \ 2.127) \cdot 10^{-3}.$$

Виконавши всі розрахунки, отримаємо, що

$$\vec{\beta}_{val}^* = \begin{pmatrix} 2.185826 \\ 0.2180424 \end{pmatrix} \cdot 10^{-4}.$$

Отже, ми отримали оцінку параметрів регресійної моделі. Зобразимо на другому рисунку пряму, яку задають значення оцінок параметрів β_0 і β_1 .

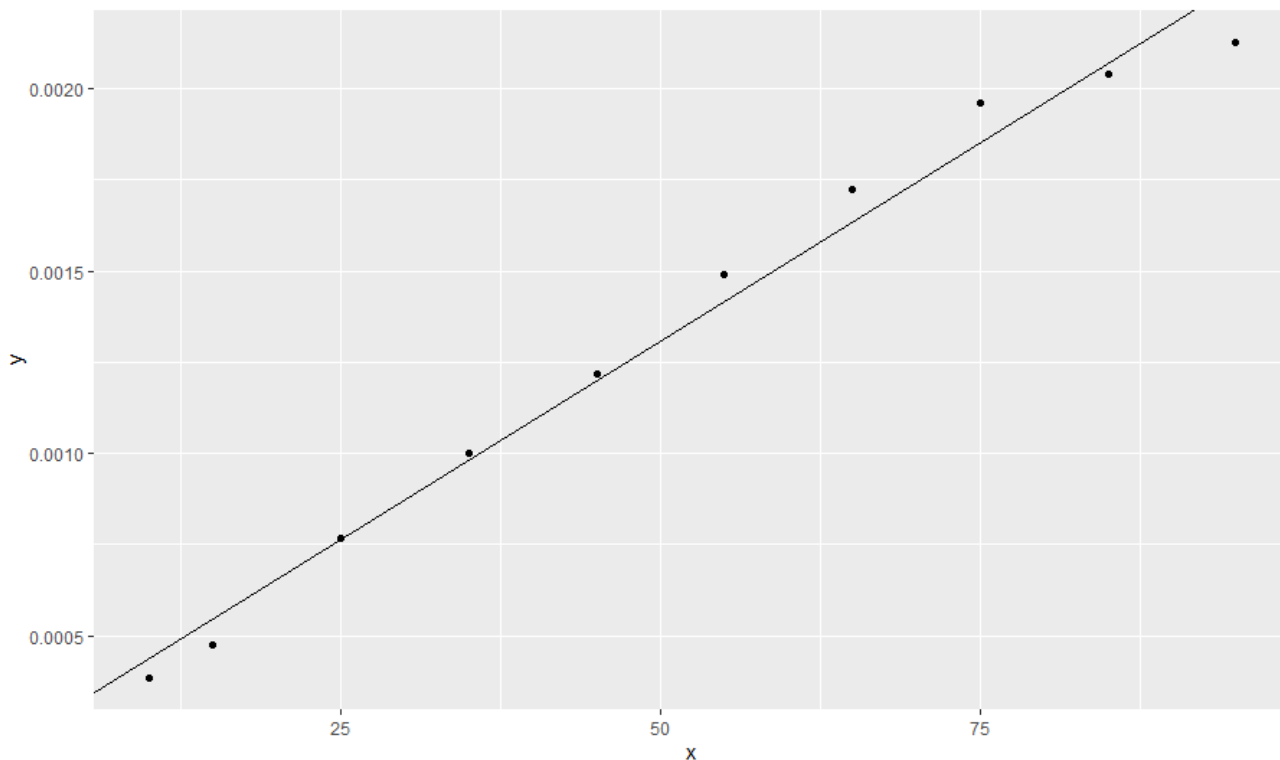


Рис. 3. Точки на площині та пряма

1.4. Перевірка адекватності побудованої моделі

Висунемо нульову гіпотезу, що константа та побудована модель не відрізняються. Альтернативною оберемо гіпотезу, що побудована модель краща за константу.

Для перевірки адекватності побудованої моделі скористаємося F -критерієм - ми хочемо порівняти залишкову оцінку дисперсії з незміщеною оцінкою дисперсії. Відомо, що статистика

$$\zeta = \frac{\frac{1}{n-1} \sum_{i=1}^n (\eta_i^{(g)} - \overline{\eta^{(g)}})^2}{\frac{1}{n-m} \sum_{i=1}^n (\eta_i^{(g)} - g^*(x_i))^2} \sim F(n-1, n-m), \quad \text{де } m - \text{кількість параметрів.}$$

Виразуємо значення, якими будемо користуватися згодом:

$$(\mathbb{D}^{**}\eta^{(g)})_{val} = \frac{1}{9} \sum_{i=1}^{10} (y_i^{(g)} - \overline{y^{(g)}})^2 = 4.198214 \cdot 10^{-7}$$

$$(\sigma_{(g)}^2)^{**}_{val} = \frac{1}{8} \sum_{i=1}^{10} (y_i^{(g)} - g^*(x_i))^2 = 7.550273 \cdot 10^{-9}$$

i

$$A^{-1} = (F^T F)^{-1} = \begin{pmatrix} 0.426 & -0.0064 \\ -0.0064 & 0.00012 \end{pmatrix}.$$

Надалі елементи матриці A^{-1} позначатимемо маленькими літерами a з індексами, якщо не вказано інше. Вирахуємо значення статистики.

$$\zeta_{val} = \frac{(\mathbb{D}^{**}\eta^{(g)})_{val}}{(\sigma_{(g)}^2)^{**}_{val}} = \frac{4.198214 \cdot 10^{-7}}{7.550273 \cdot 10^{-9}} = 55.603.$$

На рівні значущості $\alpha = 0.05$ маємо $t_{cr} = 3.39$. Оскільки критична область правостороння і $\zeta_{val} > t_{cr}$, то нульову гіпотезу відхиляємо.

Отже, побудовану модель можна вважати адекватною.

1.5. Перевірка гіпотези про значущість найменшого значення параметра побудованої моделі

Оскільки ми з'ясували, що модель можна вважати адекватною, то перевіримо на значущість параметр β_0 . Для цього висунемо нульову гіпотезу $H_0 : \beta_0 = 0$. Альтернативна гіпотеза $H_1 : \beta_0 > 0$.

Відомо, що статистика

$$\gamma^{(g)} = \frac{\beta_0^*}{\sqrt{(\sigma_{(g)}^2)^{**} \cdot a_{00}}} \sim St_{n-m},$$

за умови, що H_0 справджується.

Вирахуємо значення статистики.

$$\gamma_{val}^{(g)} = \frac{2.185826 \cdot 10^{-4}}{\sqrt{7.550273 \cdot 10^{-9} \cdot 0.426}} = 3.854.$$

На рівні значущості $\alpha = 0.05$ маємо $t_{cr} = 1.86$. Оскільки критична область правостороння і $\zeta_{val} > t_{cr}$, то нульову гіпотезу відхиляємо.

Перевіримо на значущість і параметр β_1 .

Для цього висунемо нульову гіпотезу $H_0 : \beta_1 = 0$. Альтернативна гіпотеза $H_1 : \beta_1 > 0$.

Відомо, що статистика

$$\gamma^{(g)} = \frac{\beta_1^*}{\sqrt{(\sigma_{(g)}^2)^{**} \cdot a_{11}}} \sim St_{n-m},$$

за умови, що H_0 справджується.

Вирахуємо значення статистики.

$$\gamma_{val}^{(g)} = \frac{0.2180424 \cdot 10^{-4}}{\sqrt{7.550273 \cdot 10^{-9} \cdot 0.0001}} = 25.09.$$

На рівні значущості $\alpha = 0.05$ маємо $t_{cr} = 1.86$. Оскільки критична область правостороння і $\zeta_{val} > t_{cr}$, то нульову гіпотезу відхиляємо.

Отже, наша модель є адекватною і зменшити кількість параметрів не вдалося.

1.6. Побудова прогнозованого довірчого інтервала для середнього значення відклику та самого значення відклику

Будемо будувати обидва довірчих інтервала для точки $\vec{x} = \begin{pmatrix} 1 \\ 50 \end{pmatrix}$.

Знайдемо довірчий інтервал для середнього значення відклику. Відомо, що статистика

$$\frac{g^*(x) - g(x)}{\vec{x}^T A^{-1} \vec{x}} \sim St_{n-m}.$$

Тоді довірчий інтервал для середнього значення відклику має вигляд

$$g(x) \in \left(g^*(x) - t \sqrt{(\sigma^2)_{val}^{**} \vec{x}^T A^{-1} \vec{x}}, g^*(x) + t \sqrt{(\sigma^2)_{val}^{**} \vec{x}^T A^{-1} \vec{x}} \right).$$

Вирахуємо

$$\begin{aligned} \vec{x}^T A^{-1} \vec{x} &= 0.1, \\ \sqrt{(\sigma^2)_{val}^{**} \vec{x}^T A^{-1} \vec{x}} &= 2.74 \cdot 10^{-5}, \\ g^*(50) &= \beta_0^* + \beta_1^* \cdot 50 = 12.68 \cdot 10^{-4}. \end{aligned}$$

При рівні надійності $\gamma = 0.95$ маємо $t = t_{cr} = 2.306$. Підставляючи усі знайдені значення маємо, що

$$g(x) \in (12.6 \cdot 10^{-4} - 2.306 \cdot 2.74 \cdot 10^{-5}, 12.6 \cdot 10^{-4} + 2.306 \cdot 2.74 \cdot 10^{-5}) \Leftrightarrow \\ \Leftrightarrow g(x) \in (0.00119, 0.00132).$$

А отже довірчий інтервал для середнього значення відклику $f \in (\frac{1}{0.00132}, \frac{1}{0.00119}) \Leftrightarrow (755.75, 835.55)$

Знайдемо довірчий інтервал для самого значення відклику. Відомо, що статистика

$$\frac{\eta - g^*(x)}{(\sigma^2)_g^{**}(1 + \vec{x}^T A^{-1} \vec{x})} \sim St_{n-m}.$$

Тому довірчий інтервал для самого значення відклику має вигляд

$$\eta \in \left(g^*(x) - t \sqrt{(\sigma^2)_{val}^{**}(1 + \vec{x}^T A^{-1} \vec{x})}, g^*(x) + t \sqrt{(\sigma^2)_{val}^{**}(1 + \vec{x}^T A^{-1} \vec{x})} \right).$$

Акуратно підставивши значення отримаємо

$$\eta \in (12.6 \cdot 10^{-4} - 2.306 \cdot 9.11 \cdot 10^{-5}, 12.6 \cdot 10^{-4} + 2.306 \cdot 9.11 \cdot 10^{-5}) \Leftrightarrow \\ \Leftrightarrow \eta \in (0.00104, 0.00147).$$

Повернувшись до функції f матимемо, що інтервал для самого значення відклику $\in (\frac{1}{0.00147}, \frac{1}{0.00104}) \Leftrightarrow (680.23, 952.45)$.

1.7. Висновок

На момент аналізу даних у мене були думки щодо двох моделей. Після деяких спроб модифікації даних я отримав, що якщо значення y замінити на обернені величини, то точки на графіку будуть знаходитися майже на одній прямій. Таким чином можна було розглядати обернену функцію і будувати лінійну регресійну модель відштовхуючись від цього.

Після цих думок з'явилася думка, що можна задати вигляд функції f як $f = \beta_0 + \beta_1 \cdot \frac{1}{x}$. Перевагою цією моделі від першої є те, що тут не потрібно переходити між функціями. Але я побудував графіки обох оцінок функцій - і мені здалося, що перша модель більш точно виражає поведінку даних. Саме тому модель виду $f(x) = \frac{1}{\beta_0 + \beta_1 x}$ була обрана для оцінки.

2. Задача 2

2.1. Постановка задачі

x_1	36	48	55.5	48	44.1	80	60	50	54.6	43	66	53.5	45	45	50.6
x_2	5	15	10	10	25	10	12	15	20	10	5	15	12	5	10
y	11	22.5	26	18.5	13.2	25.8	17	18	21	14.5	23	19.5	14.2	13.3	16.1

Основна мета - побудувати регресійну модель та зробити її аналіз.

2.2. Пошук оцінок параметрів двофакторної регресійної моделі за методом найменших квадратів

Будуватимемо регресійну модель виду $f(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

Зробімо припущення, що вектор похибок (похибки незалежні) має нормальний розподіл з нульовим математичним сподіванням. Використаймо метод найменших квадратів. Він полягає у знаходженні таких значень параметрів β_0 , β_1 та β_2 , щоб мінімізувати значення

$$\sum_{i=1}^{15} (\eta_i - f(\vec{x}_i))^2.$$

Відомо, що оцінкою методом найменших квадратів параметрів лінійної регресії за умови, що вектор похибок (похибки незалежні) має нормальний розподіл з нульовим математичним сподіванням, є вектор

$$\vec{\beta}^* = (F^T F)^{-1} F^T \vec{\eta}, \quad \text{де } F - \text{матриця плану.}$$

В умовах нашої задачі

$$F = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 & 1 \\ 36 & 48 & 55.5 & \dots & 45 & 45 & 50.6 \\ 5 & 15 & 10 & \dots & 12 & 5 & 10 \end{pmatrix}^T$$

і

$$\vec{\eta} = (11 \ 22.5 \ 26 \ \dots \ 14.2 \ 13.3 \ 16.1).$$

Виконавши всі розрахунки, отримаємо, що

$$\vec{\beta}_{val}^* = \begin{pmatrix} -0.2736 \\ 0.3413 \\ 0.065 \end{pmatrix}.$$

Отже, ми отримали значення оцінок параметрів регресійної моделі. Зобразимо на рисунку площину, яку задають значення оцінок параметрів β_0 , β_1 та β_2 , а також точки з умови.

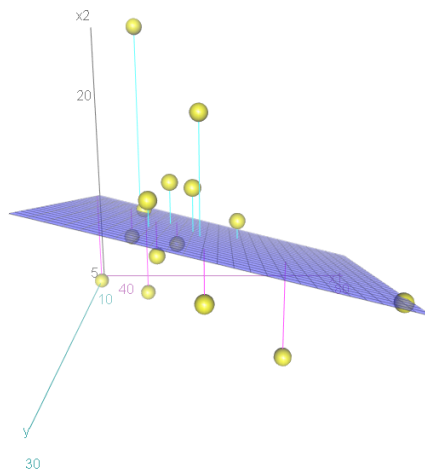


Рис. 4. Точки у просторі та площина

2.3. Перевірка адекватності побудованої моделі

Висунемо нульову гіпотезу, що константа та побудована модель не відрізняються. Альтернативною оберемо гіпотезу, що побудована модель краща за константу.

Для перевірки адекватності побудованої моделі скористаємося F -критерієм - ми хочемо порівняти залишкову оцінку дисперсії з незміщеною оцінкою дисперсії. Відомо, що статистика

$$\zeta = \frac{\frac{1}{n-1} \sum_{i=1}^n (\eta_i - \bar{\eta})^2}{\frac{1}{n-m} \sum_{i=1}^n (\eta_i - f^*(\vec{x}_i))^2} \sim F(n-1, n-m), \quad \text{де } m - \text{кількість параметрів.}$$

Вирахуємо значення, якими будемо користуватися згодом:

$$(\mathbb{D}^{**}\eta)_{val} = \frac{1}{14} \sum_{i=1}^{15} (y_i - \bar{y})^2 = 21.868$$

i

$$(\sigma^2)^{**}_{val} = \frac{1}{12} \sum_{i=1}^{15} (y_i - f^*(\vec{x}_i))^2 = 10.059$$

i

$$A^{-1} = (F^T F)^{-1} = \begin{pmatrix} 2.221 & -0.033 & -0.033 \\ -0.033 & 0.0006 & 0.00009 \\ -0.033 & 0.00009 & 0.0023 \end{pmatrix}.$$

Надалі елементи матриці A^{-1} позначатимемо маленькими літерами a з індексами, якщо не вказано інше. Вирахуємо значення статистики.

$$\zeta_{val} = \frac{(\mathbb{D}^{**}\eta)_{val}}{(\sigma^2)^{**}_{val}} = \frac{21.868}{10.059} = 2.17.$$

На рівні значущості $\alpha = 0.05$ маємо $t_{cr} = 2.64$. Оскільки критична область правостороння і $\zeta_{val} < t_{cr}$, то нульову гіпотезу не відхиляємо.

Отже, побудовану модель не можна вважати адекватною на рівні значущості $\alpha = 0.05$. Це означає, що, можливо, варто спробувати інший вигляд моделі. Але цього ми робити не будемо. Детально про пошук іншої моделі буде написано у висновку.

Проте варто зауважити, що на рівні значущості $\alpha = 0.25$ значення $t_{cr} = 2.11$, а тому модель можна вважати адекватною на рівні значущості $\alpha = 0.25$.

Отже, модель є не такою вже й поганою, а тому ми продовжимо з нею працювати.

2.4. Перевірка гіпотези про значущість найменшого значення параметра побудованої моделі

Оскільки ми з'ясували, що модель можна вважати адекватною на певному рівні значущості, то перевіримо на значущість параметр β_2 . Для цього висунемо нульову гіпотезу $H_0 : \beta_2 = 0$. Альтернативна гіпотеза $H_1 : \beta_2 > 0$.

Відомо, що статистика

$$\gamma = \frac{\beta_2^*}{\sqrt{(\sigma^2)^{**} \cdot a_{22}}} \sim St_{n-m},$$

за умови, що H_0 справджується.

Вирахуємо значення статистики.

$$\gamma_{val} = \frac{0.065}{\sqrt{10.059 \cdot 0.0023}} = 0.42.$$

На рівні значущості $\alpha = 0.05$ маємо $t_{cr} = 1.782$. Оскільки критична область правостороння і $\zeta_{val} < t_{cr}$, то нульову гіпотезу не відхиляємо.

Це означає, що другий фактор не вносить вагомого внеску до результату побудованої моделі. Тому модель можна спростити до вигляду $g(x_1) = \beta_0 + \beta_1 x_1$.

Обрахуємо значення

$$(\sigma_{(g)}^2)^{**}_{val} = \frac{1}{12} \sum_{i=1}^{15} (y_i - g^*(x_{1i}))^2 = 10.9633.$$

Порівнюючи значення оцінки з критичним значенням отримуємо, що модель не є адекватною ні на рівні значущості $\alpha = 0.05$, ні на рівні значущості $\alpha = 0.25$.

Отож, на цьому етапі можна лишити усі намагання витягти з моделі хоча б щось, адже побудована модель за F -критерієм залишатиметься неадекватною. Але це нам допомогло зрозуміти, що, можливо, варто будувати однофакторну модель.

2.5. Побудова нової однофакторної моделі

Для цього зобразимо точки виду (x_1, y) на площині.

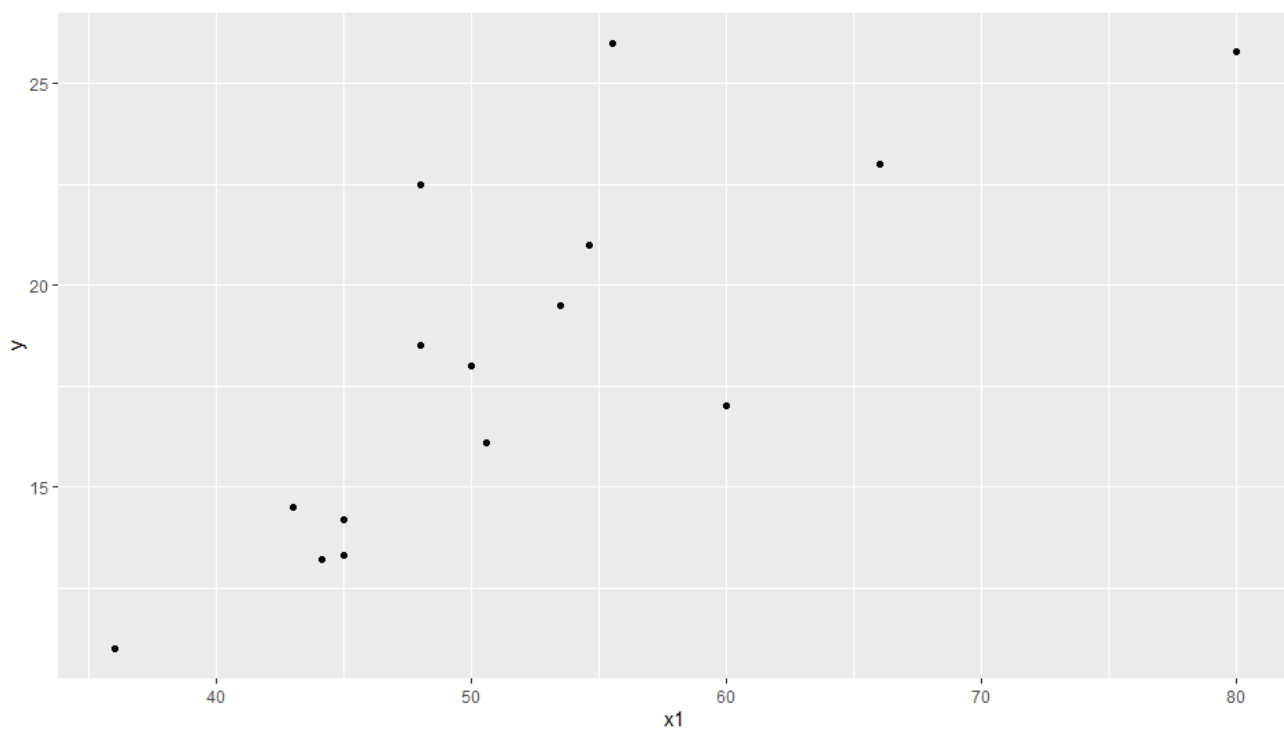


Рис. 5. Залежність між x_1 та y

Зобразимо тепер залежність між $\ln(x_1)$ та y .

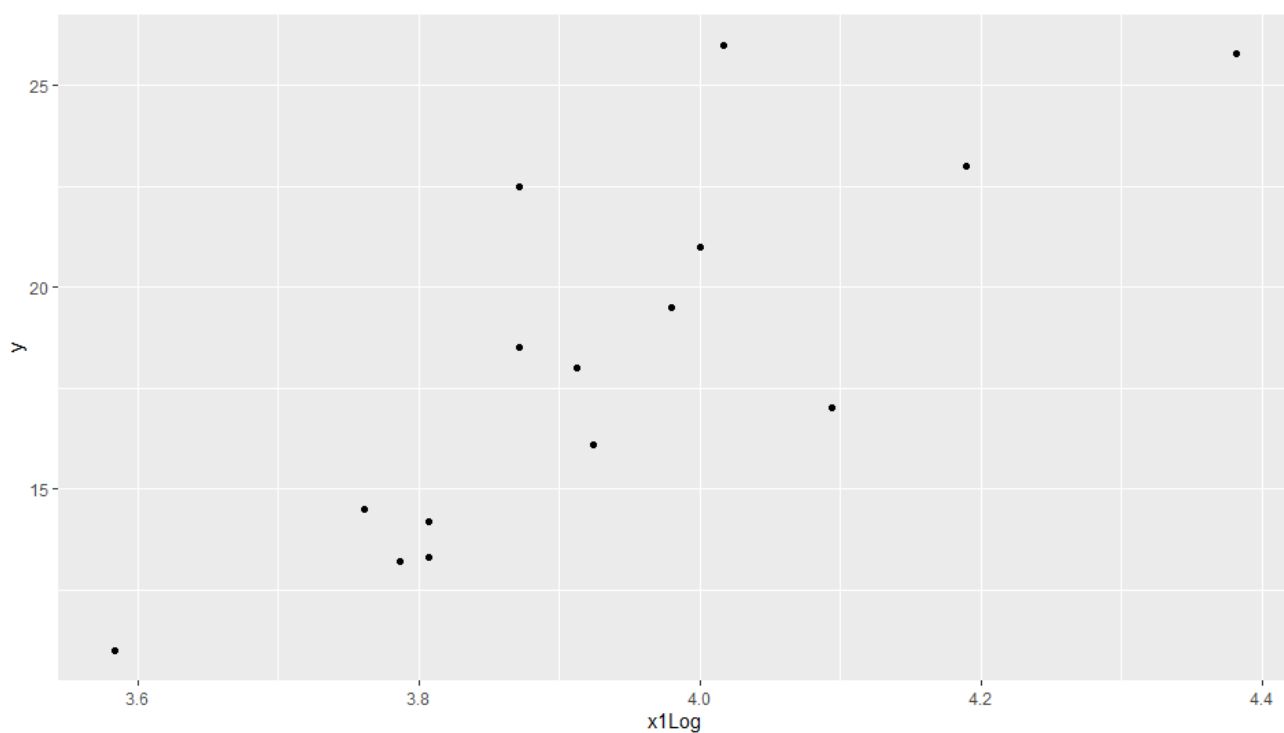


Рис. 6. Залежність між $\ln(x_1)$ та y

Бачимо, що залежність схожа на лінійну. Давайце це перевіримо. Розглядатимемо модель виду $f(x) = \beta_0 + \beta_1 \ln(x)$. Знайдемо оцінку параметрів.

В умовах нашої задачі

$$F = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 & 1 \\ 3.58 & 3.87 & 4.016 & \dots & 3.8 & 3.8 & 3.92 \end{pmatrix}^T$$

і

$$\vec{\eta} = (11 \quad 22.5 \quad 26 \quad \dots \quad 14.2 \quad 13.3 \quad 16.1).$$

Виконавши всі розрахунки, отримаємо, що

$$\vec{\beta}_{val}^* = \begin{pmatrix} -58.071 \\ 19.406 \end{pmatrix}.$$

Зобразимо на графіку дану пряму.

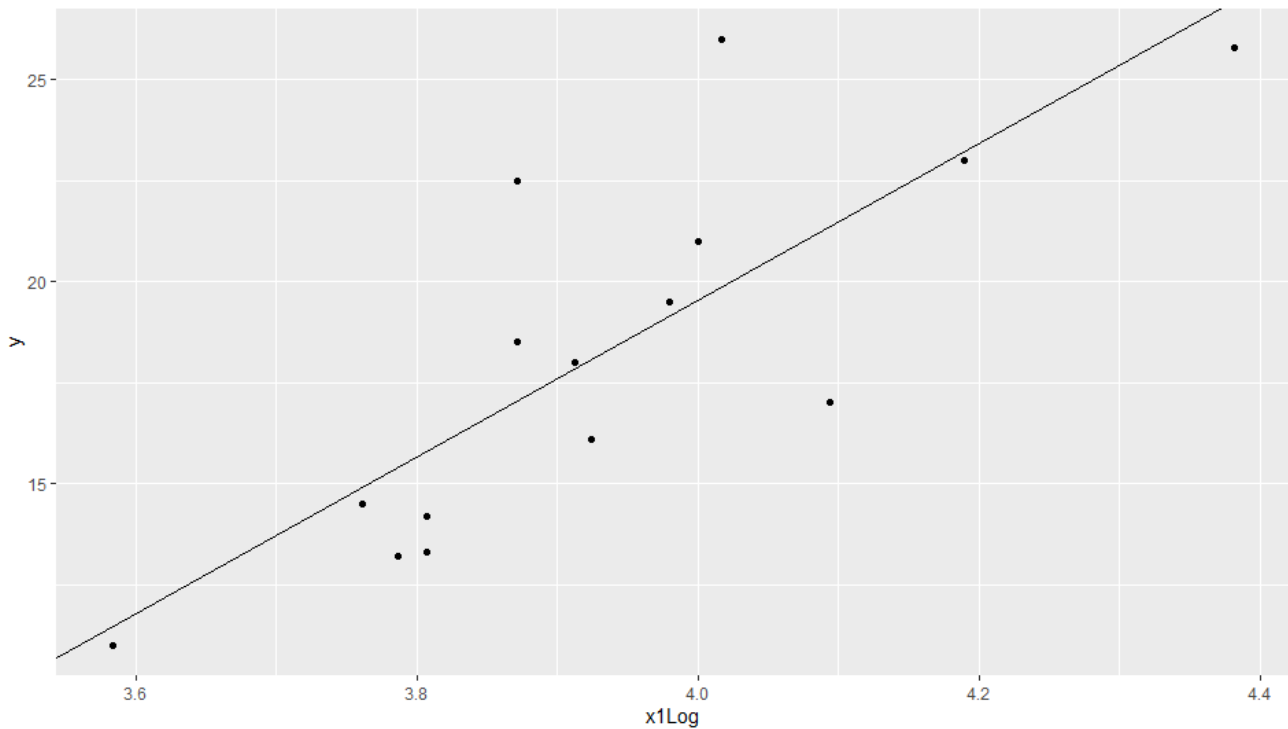


Рис. 7. Залежність між $\ln(x_1)$ та y і пряма

2.6. Перевірка адекватності нової побудованої однофакторної моделі

Висунемо нульову гіпотезу, що константа та побудована модель не відрізняються. Альтернативною оберемо гіпотезу, що побудована модель краща за константу.

Для перевірки адекватності побудованої моделі скористаємося F -критерієм - ми хочемо порівняти залишкову оцінку дисперсії з незміщеною оцінкою дисперсії. Відомо, що статистика

$$\zeta = \frac{\frac{1}{n-1} \sum_{i=1}^n (\eta_i - \bar{\eta})^2}{\frac{1}{n-m} \sum_{i=1}^n (\eta_i - f^*(x_i))^2} \sim F(n-1, n-m), \quad \text{де } m - \text{кількість параметрів.}$$

Вирахуємо значення, якими будемо користуватися згодом:

$$(\mathbb{D}^{**}\eta)_{val} = \frac{1}{14} \sum_{i=1}^{15} (y_i - \bar{y})^2 = 21.868$$

i

$$(\sigma^2)^{**}_{val} = \frac{1}{13} \sum_{i=1}^{15} (y_i - f^*(x_i))^2 = 8.4763$$

i

$$A^{-1} = (F^T F)^{-1} = \begin{pmatrix} 29.7833 & -7.557 \\ -7.557 & 1.921 \end{pmatrix}.$$

Надалі елементи матриці A^{-1} позначатимемо маленькими літерами a з індексами, якщо не вказано інше. Вирахуємо значення статистики.

$$\zeta_{val} = \frac{(\mathbb{D}^{**}\eta)_{val}}{(\sigma^2)^{**}_{val}} = \frac{21.868}{8.4763} = 2.579.$$

На рівні значущості $\alpha = 0.05$ маємо $t_{cr} = 2.55$. Оскільки критична область правостороння і $\zeta_{val} > t_{cr}$, то нульову гіпотезу відхиляємо.

Отже, побудовану модель можна вважати адекватною на рівні значущості $\alpha = 0.05$.

2.7. Перевірка гіпотези про значущість найменшого значення параметра нової побудованої однофакторної моделі

Оскільки ми з'ясували, що модель можна вважати адекватною, то перевіримо на значущість параметра β_1 . Для цього висунемо нульову гіпотезу $H_0 : \beta_1 = 0$. Альтернативна гіпотеза $H_1 : \beta_1 > 0$.

Відомо, що статистика

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

за умови, що H_0 справджується.

Вирахуємо значення статистики.

$$\gamma_{val} = \frac{19.406}{\sqrt{8.4763 \cdot 1.9217}} = 4.808.$$

На рівні значущості $\alpha = 0.05$ маємо $t_{cr} = 1.771$. Оскільки критична область правостороння і $\zeta_{val} > t_{cr}$, то нульову гіпотезу відхиляємо.

Перевіримо на значущість і параметр β_0 . Для цього висунемо нульову гіпотезу $H_0 : \beta_0 = 0$. Альтернативна гіпотеза $H_1 : \beta_0 < 0$.

Відомо, що статистика

$$\gamma = \frac{\beta_0^*}{\sqrt{(\sigma^2)^{**} \cdot a_{00}}} \sim St_{n-m}.$$

, за умови, що H_0 справджується.

Вирахуємо значення статистики.

$$\gamma_{val} = \frac{-58.071}{\sqrt{8.4763 \cdot 29.7833}} = -3.65.$$

На рівні значущості $\alpha = 0.05$ маємо $t_{cr} = -1.771$. Оскільки критична область лівостороння і $\zeta_{val} < t_{cr}$, то нульову гіпотезу відхиляємо.

Отже, наша модель є адекватною і зменшити кількість параметрів не вдалося.

2.8. Побудова прогнозованого довірчого інтервала для середнього значення відклику та самого значення відклику нової однофакторної моделі

Будемо будувати обидва довірчих інтервала для точки $\vec{x} = \begin{pmatrix} 1 \\ \ln(52) \end{pmatrix}$.

Знайдемо довірчий інтервал для середнього значення відклику. Відомо, що статистика

$$\frac{f^*(x) - f(x)}{\vec{x}^T A^{-1} \vec{x}} \sim St_{n-m}.$$

Тоді довірчий інтервал для середнього значення відклику має вигляд

$$f(x) \in \left(f^*(x) - t\sqrt{(\sigma^2)_{val}^{**} \vec{x}^T A^{-1} \vec{x}}, f^*(x) + t\sqrt{(\sigma^2)_{val}^{**} \vec{x}^T A^{-1} \vec{x}} \right).$$

Виразуємо

$$\begin{aligned} \vec{x}^T A^{-1} \vec{x} &= 4.089, \\ \sqrt{(\sigma^2)_{val}^{**} \vec{x}^T A^{-1} \vec{x}} &= 5.887, \\ f^*(52) &= \beta_0^* + \beta_1^* \cdot \ln(52) = 18.6. \end{aligned}$$

При рівні надійності $\gamma = 0.95$ маємо $t = t_{cr} = 2.16$. Підставляючи усі знайдені значення маємо, що

$$f(x) \in (18.6 - 2.16 \cdot 5.887, 18.6 + 2.16 \cdot 5.887) \Leftrightarrow f(x) \in (5.88, 31.31).$$

Знайдемо довірчий інтервал для самого значення відклику. Відомо, що статистика

$$\frac{\eta - f^*(x)}{(\sigma^2)^{**}(1 + \vec{x}^T A^{-1} \vec{x})} \sim St_{n-m}.$$

Тому довірчий інтервал для самого значення відклику має вигляд

$$\eta \in \left(f^*(x) - t\sqrt{(\sigma^2)_{val}^{**}(1 + \vec{x}^T A^{-1} \vec{x})}, f^*(x) + t\sqrt{(\sigma^2)_{val}^{**}(1 + \vec{x}^T A^{-1} \vec{x})} \right).$$

Акуратно підставивши значення отримаємо

$$\eta \in (18.6 - 2.16 \cdot 6.56, 18.6 + 2.16 \cdot 6.56) \Leftrightarrow \eta \in (4.43, 32.77).$$

2.9. Висновок

Спочатку для дослідження була обрана найпростіша двофактора лінійна модель, але вона не пройшла тест на адекватність. Потім з'ясувалося, що деякий фактор виявився незначущим. Усі ці чинники змусили мене шукати інший вигляд моделі.

Я перебрав на комп'ютері усі моделі виду $f(x_1, x_2) = \beta_0 + \beta_1 x_1^i + \beta_2 x_2^j$, для всіх цілих $i, j \in [-50, 50]$. Серед них не було жодної, що б пройшла тест на адекватність, хоча були ті, які були доволі близько до критичного значення. Але для великих i, j потрібно було досліджувати похибки, тому ніяка модель такого вигляду мене не влаштовувала.

Далі була ідея додати фактор x_2^2 та $\sqrt{x_1}$, але побудована модель знов таки не пройшла тест на адекватність.

Після аналізу усіх побудованих моделей виникло припущення, що фактор x_2 не вносить вагомого внеску до більшості моделей. Тому було прийнято рішення почати пошук однофакторних моделей виду $f(x) = \beta_0 + \beta_1 x^i$, для деякого цілого i . Були адекватні моделі, а з доволі малим значенням i . Але б їх було занадто складно досліджувати в плані похибок. Бо навіть для вирахування матриці A^{-1} доволі малі зміни в числах матриці A приводять до іншого вигляду оберненої матриці. Потім стало зрозуміло, що шукати вигляд базисної функції варто серед експоненціальних функцій. Найпершою було обрано функцію $\ln(x)$, з якою модель пройшла тест на адекватність та підтвердила значущість параметрів.

Література

- [1] Kaniovska I. Y. *Конспект лекцій з теорії ймовірностей та математичної статистики.*