

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
«ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ»
Кафедра прикладної математики та інформатики

МЕТОДИЧНІ ВКАЗІВКИ

ДО ВИКОНАННЯ ПРАКТИЧНОЇ ТА САМОСТІЙНОЇ РОБОТИ
з дисципліни «Теорія та практика комп'ютерних експериментів» для
аспірантів спеціальності 122 Комп'ютерні науки всіх форм навчання

УДК 004.94:519.8
М54

Методичні вказівки до виконання практичної та самостійної роботи з дисципліни «Теорія та практика комп'ютерних експериментів» для аспірантів спеціальності 122 Комп'ютерні науки всіх форм навчання [Електронний ресурс] / укл. : О.М. Любименко, А. О. Нікітенко. – Луцьк : ДонНТУ, 2023. – 95 с.

У методичних вказівках надано рекомендації до виконання практичних робіт аспірантами при вивченні дисципліни «Теорія та практика комп'ютерних експериментів».

Наведено приклади розв'язування завдань, які розраховуються при виконанні практичних робіт, наведено вихідні дані практичних завдань, питання для самостійної роботи, перелік рекомендованої літератури.

Методичні вказівки можуть бути використані аспірантами денної та заочної форми навчання спеціальності 122 Комп'ютерні науки.

Укладачі: Любименко О.М., доц., к.ф.-м.н., доц. кафедри прикладної математики та інформатики;
Нікітенко А.О., асистент кафедри прикладної математики та інформатики

Рецензент: Штепа О.А. доц., к.т.н., доц. кафедри електронної техніки

Відповідальний за випуск: Маслова Н.О., доц., к.т.н., в.о. завідувача кафедри прикладної математики та інформатики

Затверджено навчально-методичним відділом ДонНТУ,
Протокол № 7 від 25.04.2023 р.

Розглянуто на засіданні кафедри прикладної математики та інформатики,
протокол № 4_ від 18.04. 2023 року

ЗМІСТ

ВСТУП	5
1 ЗАСОБИ ДІАГНОСТИКИ РЕЗУЛЬТАТІВ НАВЧАННЯ	9
2 ЗАВДАННЯ ТА РЕКОМЕНДАЦІЇ ДО ВИКОНАННЯ ПРАКТИЧНИХ РОБІТ	11
2.1 ПРАКТИЧНА РОБОТА № 1 - КРИТЕРІЇ ОПТИМАЛЬНОСТІ ПЛАНІВ ЕКСПЕРИМЕНТУ	11
Основні теоретичні положення.	11
Приклади розв'язання	16
Завдання для виконання на практичному занятті	25
Питання для самостійної роботи	28
Рекомендована література	28
2.2 ПРАКТИЧНА РОБОТА №2 - ПОВНИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ.	29
Основні теоретичні питання	29
Приклади розв'язання	32
Завдання для виконання на практичному занятті	48
Питання для самостійної роботи	55
Рекомендована література	55
2.3 ПРАКТИЧНА РОБОТА №3 - ДРОБОВИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ	56
Основні теоретичні положення.	56
Приклади розв'язання	58
Завдання для виконання на практичному занятті	68
Питання для самостійної роботи	73
Рекомендована література	73
2.4 ПРАКТИЧНА РОБОТА №4 - ОБРОБКА РЕЗУЛЬТАТІВ ПРАКТИЧНОГО ЕКСПЕРИМЕНТУ	74

Основні теоретичні положення.	74
Приклади розв'язання	81
Завдання для виконання на практичному занятті	86
Питання для самостійної роботи	87
Рекомендована література	88
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ	89
ДОДАТОК А. КРИТЕРІЙ СТЬЮДЕНТА.	91
ДОДАТОК Б. КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ КОХРЕНА	93
ДОДАТОК В. КРИТЕРІЙ ФІШЕРА	94
ДОДАТОК Г. КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ χ^2	95

ВСТУП

Планувати експеримент людину змусило саме життя. Стародавня людина при полюванні на великого звіра повинна була вибрати оптимальну масу і форму каменю, з яким він вирушав за видобутком в залежності від багатьох факторів: розміру планованого видобутку, своїх фізичних можливостей, відстані і т. д. В результаті експериментальної перевірки з'являлися досконаліші знаряддя для полювання, землеробства та війни. Природно, що статистичні методи планування експерименту з'явилися значно пізніше – на початку XX століття. Їх основи були закладені в роботах англійського статистика Рональда Фішера, в яких автор довів доцільність одночасного варіювання всіма факторами, що істотно впливають на параметр, що оптимізується, на відміну від широко поширеного до цього однофакторного експерименту. Найбільш ефективне практичне застосування планування екстремального експерименту, теоретичні основи якого заклали G. Box та K. Wilson, отримало в 40-60 роках XX століття при реалізації Манхеттенського проекту в США. Найскладніші теоретичні гіпотези були перевірені в ході масштабних експериментальних досліджень, які потрібно провести в найкоротший термін. Ідея методу Бокса–Уилсона у тому, що ставилися невеликі серії дослідів, у кожному з яких одночасно варіювалися за певними правилами чинники. Серії організовувалися таким чином, щоб за результатами попередньої серії можна було спланувати умови проведення подальших дослідів. Таким чином, покроково досягався запланований результат щодо знаходження області оптимуму. Розвиток теорії та практики планування експерименту отримало у роботах Хантера, Д. Фінта, Дж. Кіфера, Х. Шенка, Д.К. Монтгомері, В.В. Налімова Ю.П. Адлера, І.П. Копилова та інших вчених.

Таким чином, для успішного вирішення наукових, виробничих, і технологічних проблем і завдань, інженеру досліднику необхідно мати компетенції в галузі математичної теорії та практичних навичок планування експерименту та методів обробки експериментальних даних.

Методи планування експерименту застосовуються для побудови інтерполяційних моделей та оптимізації процесів та об'єктів. У методичних вказівках розглядаються умови та необхідність застосування математичної теорії планування експерименту у наукових та інженерних дослідженнях. Алгоритми побудови планів експерименту можна застосувати до більшості оптимізаційних завдань як проектно-розрахункових, так і експериментальних. Методи статистичної оцінки досвідчених даних необхідні будь-якого експериментального дослідження, претендує достовірність результатів.

Метою викладання навчальної дисципліни «Теорія та практика комп'ютерних експериментів» є засвоєння аспірантом фундаментальних понять планування й організації комп'ютерних експериментів, методів отримання інформації під час експерименту, оптимальної організації як наукового, так і інженерного експерименту, статистичної обробки результатів експерименту, а також отримання знань і практичних навичок у галузі методів і засобів застосування сучасних інформаційних технологій і обладнання для автоматизації експериментальних досліджень.

Успішному вивченню дисципліни «Теорія та практика комп'ютерних експериментів» сприяє попереднє опанування такими дисциплінами, як «Сучасне наукове дослідження: організація, виконання, результат», «Філософія та наукове пізнання».

Основною метою методичних вказівок з навчальної дисципліни «Теорія та практика комп'ютерних експериментів» є набуття практичних навичок, які надають можливість самостійно застосовувати методи

виключення результатів з грубими похибками; способи перевірки відповідності нормальному розподілу отриманих даних; робастні методи обробки даних; непараметричні методи обробки даних; будувати плани багатофакторного експерименту.

У результаті вивчення навчальної дисципліни аспірант повинен знати:

Основними результатами опанування дисципліною «Теорія та практика комп'ютерних експериментів» є:

- розвиток творчих здібностей щодо прогнозування напрямків та шляхів розвитку науки та практики дослідження програмного забезпечення інформаційно-комп'ютерних систем з використанням сучасних підходів щодо виконання експериментів.
- глибоке та творче засвоєння теоретичних знань щодо планування, проведення та обробки експериментів з інформаційно-комп'ютерними системами.
- стимулювання здобувачів до інноваційної діяльності у галузі інформаційних технологій.
- сприяння теоретичній підготовці здобувачів для вирішення завдань дисертаційних досліджень у галузі інформаційних технологій з використанням експериментів.
- активізація творчого відношення до професійної діяльності
- застосовування отриманих теоретичних знань на практиці та аналіз отриманих результатів.

Внаслідок вивчення курсу студенти повинні вміти:

- застосовувати знання для постановки і вирішення актуальних наукових завдань, обґрунтування, та використання відповідних методів проведення експериментів з інтелектуальними інформаційними системами.

- проектувати плани проведення експериментів та обирати засоби проведення експериментів з інтелектуальними інформаційними системами.
- обґрунтовувати обрані способи обробки і аналізу результатів експериментів з інтелектуальними інформаційними системами.
- застосовувати принципи навчання впродовж життя для вирішення завдань поза програмою курсу;
- застосовувати сучасні математичні програмні пакети для обробки результатів експерименту.

1 ЗАСОБИ ДІАГНОСТИКИ РЕЗУЛЬТАТІВ НАВЧАННЯ

Засобами оцінювання та методами демонстрування результатів навчання при опануванні дисципліною «Теорія та практика комп'ютерних експериментів»:

- залік;
- індивідуальні завдання до виконання практичних робіт;
- презентації результатів виконаних завдань та досліджень.

Критерії оцінювання результатів навчання

Критерії оцінювання мають формулювати порядок оцінювання під час поточного контролю за результатами практичних занять, та підсумкового контролю.

Форма навчання	Вид роботи				Поточний контроль	Залік
	Пр 1	Пр 2	Пр 3	Пр 4		
Денна, заочна	25	25	25	25	100	100

Оцінювання знань студента здійснюється за 100-бальною шкалою. При оформленні документів за екзаменаційну сесію використовується таблиця відповідності оцінювання знань студентів за наступною шкалою:

Оцінка		
За 100-бальною шкалою	Для екзамену, курсового проекту(роботи), практики, диференційованого заліку, кваліфікаційного екзамену, випускної кваліфікаційної (дипломної) роботи (проекту)	Для заліку
90-100	Відмінно	зараховано
74-89	Добре	зараховано
60-73	Задовільно	зараховано
0-59	Незадовільно	не зараховано

Бали отримані при виконанні практичних робіт, або під час проведення практичних занять, враховуються при виставленні заліку. При отриманні незадовільної оцінки з будь-якого виду занять студенту назначається додаткова консультація (не більше двох на одне заняття) для перездачі.

2 ЗАВДАННЯ ТА РЕКОМЕНДАЦІЇ ДО ВИКОНАННЯ ПРАКТИЧНИХ РОБІТ

2.1 ПРАКТИЧНА РОБОТА № 1 - КРИТЕРІЇ ОПТИМАЛЬНОСТІ ПЛАНІВ ЕКСПЕРИМЕНТУ

Основні теоретичні положення.

Мета роботи: Вивчити поняття оптимального плану експерименту та критерії оптимальності планів.

Припустимо, що досліджувана модель має вигляд:

$$y = f^T(x)\theta + e = \sum_{l=1}^m f_l(x)\theta_l + e,$$

де y – значення залежної змінної; $f^T(x) = (f_1(x), \dots, f_m(x))$ – задана вектор-функція від незалежної векторної змінної x , яка може змінюватися в області $\tilde{X}$;

$\theta = (\theta_1, \dots, \theta_m)^T$ – вектор невідомих параметрів, які необхідно визначити за результатами експериментів (вимірювань);

e – помилка.

Припустимо також, що будь-яке число вимірювань залежної змінної можна провести за довільних $x_i \in \tilde{X}$. Нехай помилки вимірювань є незалежними випадковими величинами з нульовим математичним очікуванням і однаковою дисперсією. Для знаходження оцінок вектора параметрів θ використовується метод найменших квадратів.

ВИЗНАЧЕННЯ 1. Планом експерименту називається сукупність величин виду

$$\xi_N = \begin{Bmatrix} x_1, & \dots, & x_n \\ r_1 & \dots, & r_n \end{Bmatrix},$$

де $\sum_{i=1}^n r_i = N$, x_i – точка, в якій проводиться r_i спостережень:

N – загальне число спостережень.

Сукупність точок $x_1, \dots, x_n$ називається спектром плану ξ_N .

ВИЗНАЧЕННЯ 2. Дискретним нормованим планом називається сукупність величин виду

$$\varepsilon_N = \begin{Bmatrix} x_1, & \dots, & x_n \\ p_1 & \dots, & p_n \end{Bmatrix},$$

де $\sum_{i=1}^n p_i = 1, p_i \geq 0$.

ВИЗНАЧЕННЯ 3. Безперервним нормованим планом називається сукупність величин виду

$$\varepsilon = \begin{Bmatrix} x_1, & \dots, & x_n \\ p_1 & \dots, & p_n \end{Bmatrix},$$

де $\sum_{i=1}^n p_i = 1, p_i = r_i/N$.

ВИЗНАЧЕННЯ 4. Нормованою інформаційною матрицею (Фішера) дискретного або безперервного плану називається величина

$$M = \sum_{j=1}^n p_j f(x_j) f^T(x_j).$$

Щоб підкреслити залежність інформаційної матриці від плану, використовують позначення $M(\varepsilon_N)$ або $M(\varepsilon)$. Матриця, зворотна до інформаційної, називається дисперсійною, тобто.

$$D(\varepsilon) = M^{-1}(\varepsilon), D(\varepsilon_N) = M^{-1}(\varepsilon_N),$$

Безперервний нормований план у загальному випадку може відповідати ймовірнісній мірі $P_\varepsilon(x)$, заданої на області $\tilde{X}$ такою, що задовольняє умовам

$$\int_{\tilde{X}} dP_{\varepsilon}(x) = 1; P_{\varepsilon}(x) \geq 0; \forall x \in \tilde{X}$$

При цьому нормована інформаційна матриця плану ε визначається співвідношенням

$$M(\varepsilon) = \int_{\tilde{X}} f(x)f^T(x)dP_{\varepsilon}(x)$$

Перелічимо властивості інформаційної матриці

1) Для будь-якого плану ε інформаційна матриця $M(\varepsilon)$ - позитивно напіввизначена.

2) Матриця $M(\varepsilon)$ - особлива (тобто $|M(\varepsilon)| = 0$), якщо спектр плану ε містить менше ніж m точок.

3) Множина матриць $M(\varepsilon)$, що відповідає всім можливим нормованим планам, є опуклим.

4) Для будь-якого плану ε завжди знайдеться план спектр якого містить не більше ніж $\frac{m(m+1)}{2} + 1$ точок та інформаційна матриця якого $M(\tilde{\varepsilon})$, збігається з інформаційною матрицею $M(\varepsilon)$ плану ε .

Під оптимальним плануванням експерименту розумітимемо вибір плану експерименту відповідно до тих чи інших критеріїв оптимальності. Будемо оцінювати якість плану ε за значенням деякого функціонала Ψ від інформаційної матриці $M(\varepsilon)$ або відповідної їй дисперсійної матриці $D(\varepsilon)$.

ВИЗНАЧЕННЯ 5. План ε^* зветься Ψ -оптимальним, якщо

$$\varepsilon^* = \text{Arg } m_{\varepsilon} \Psi[M(\varepsilon)].$$

Основні критерії оптимальності можна розбити на дві групи: критерії, пов'язані з точністю оцінювання параметрів моделі, і критерії, пов'язані з точністю оцінювання математичного очікування функції відгуку.

Розглянемо основні найбільш відомі критерії оптимальності першої групи.

План ε^* зветься **D-оптимальним**, якщо

$$\varepsilon^* = \text{Arg max}_{\varepsilon} |M(\varepsilon)|.$$

або

$$\varepsilon^* = \text{Arg min}_{\varepsilon} |D(\varepsilon)|.$$

Еліпсоїд ε^* розсіювання оцінок параметрів для D –оптимального плану має мінімальний обсяг.

План називається **A-оптимальним**, якщо його дисперсійна матриця має найменший слід, тобто

$$\varepsilon^* = \text{Arg min}_{\varepsilon} \text{tr}D(\varepsilon).$$

A-оптимальний план дає змогу знайти оцінки невідомих параметрів із мінімальною середньою дисперсією. При цьому еліпсоїд розсіювання має мінімальну суму квадратів довжин осей і найменшу довжину діагоналі паралелепіпеда, описаного біля цього еліпсоїда.

План ε^* називається **E-оптимальним**, якщо він мінімізує (максимізує) максимальне (мінімальне) власне значення дисперсійної (інформаційної) матриці, тобто якщо

$$\varepsilon^* = \text{Arg min}_{\varepsilon} \max_i \lambda_i [D(\varepsilon)].$$

або

$$\varepsilon^* = \text{Arg min}_{\varepsilon} \max_i \lambda_i [M(\varepsilon)],$$

де λ_i - власне значення матриці $D(\varepsilon)$ або $M(\varepsilon)$.

План ε^* зветься Φ_p -**оптимальним**, якщо

$$\varepsilon^* = \text{Arg min}_{\varepsilon} \Phi_p(\varepsilon) = \text{Arg min}_{\varepsilon} (m^{-1} \text{tr} D^p(\varepsilon))^{1/p}$$

Клас функціоналів $\Phi_p(\varepsilon)$ включає щойно введені критерії оптимальності: $\Phi_0(\varepsilon)$ - відповідає критерію D-оптимальності $\Phi_1(\varepsilon)$ - критерію A-оптимальності, $\Phi_{\infty}(\varepsilon)$ - критерію E-оптимальності. У практичних ситуаціях для $0 < p < \infty$ можна використовувати замість $\Phi_p(\varepsilon)$ його еквівалент $\tilde{\Phi}_p(\varepsilon) = p^{-1} \text{tr} D^p(\varepsilon)$.

План ε^* зветься Λ - **оптимальним**, якщо

$$\varepsilon^* = \text{Arg min}_{\varepsilon} \sum_{i=1}^m [\lambda_i(D(\varepsilon)) - \bar{\lambda}(D(\varepsilon))]^2,$$

де $\bar{\lambda}$ -середнє значення власних чисел.

План ε^* зветься MV- **оптимальним**, якщо

$$\varepsilon^* = \text{Arg min}_{\varepsilon} \max_i D_{ii}(\varepsilon),$$

де $D_{ii}(\varepsilon)$ -діагональний елемент дисперсійної матриці.

Серед критеріїв другої групи виділимо критерії G-, Q-оптимальності та екстраполяції в точку.

План ε^* зветься G- **оптимальним**, якщо

$$\varepsilon^* = \text{Arg min}_{\varepsilon} \max_{x \in \tilde{X}} d(x, \varepsilon),$$

де $d(x, \varepsilon) = f^T(x)M^{-1}(\varepsilon)f(x)$ -дисперсія оцінки функції відгуку.

Використання цього критерію гарантує відсутність точок, що мають занадто низьку точність оцінювання функції відгуку.

План ε^* зветься **Q-оптимальним**, якщо

$$\varepsilon^* = \text{Arg min}_{\varepsilon} \int_Z d(x, \varepsilon) dx = \text{Arg min}_{\varepsilon} Q[M^{-1}(\varepsilon)],$$

де область Z може не збігатися з областю $\tilde{X}$.

Q-оптимальні плани мінімізують середнє по області Z значення дисперсії оцінки поверхні відгуку.

План ε^* зветься **Q-оптимальним при екстраполяції в точку x^0** , якщо

$$\varepsilon^* = \text{Arg min}_{\varepsilon} d(x_0, \varepsilon).$$

Ці плани мінімізують дисперсію $d(x_0, \varepsilon)$ оцінки функції відгуку в точці $x = x_0$.

План називається ротатабельним, якщо дисперсія оцінки функції відгуку може бути представлена як функція відстані від центру експерименту. Критерій уніформності плану вимагає, щоб дисперсія оцінки функції відгуку в деякій області в центрі експерименту була практично постійною.

Приклади розв'язання

Приклад 1. Побудувати аналітичну модель оптимальних планів.

Розв'язання

Нехай модель регресії має вигляд

$$\eta(x, \theta) = \theta_1 + \theta_2 x^2, \quad -1 \leq x \leq 1.$$

Необхідно побудувати D-оптимальний план. Припустимо, що оптимальний план належить класу планів такого виду:

$$\varepsilon = \begin{cases} -1 & 0, & 1 \\ q & 1 - 2q & q \end{cases}.$$

Знайдемо оптимальне значення параметра q .

$$|M(\varepsilon)| = 2q - 4q^2, \quad \frac{\partial |M(\varepsilon)|}{\partial q} = 2 - 8q = 0, \quad q^* = 1/4$$

$$\varepsilon = \begin{cases} -1 & 0, & 1 \\ 0,25 & 0,5 & 0,25 \end{cases}.$$

Необхідно перевірити, що отриманий план справді є D -оптимальним.

Для цього перевіримо, що $\max_x d(x, \varepsilon^*) = m = 2$.

$$\text{Дійсно } d(x, \varepsilon^*) = 2 - 4x^2 + 4x^4,$$

максимум дорівнює 2 і досягається на межах і в точці 0.

Приклад 2.

Модель квадратична на відрізку. Плани для аналізу: 1-4. (табл.2.1)

Для пункта 4 завдання використаний критерій D -оптимальності.

Таблиця 2.1- Плани для аналізу

№ п/п	x_1/p_1	x_2/p_2	x_3/p_3	x_4/p_4	x_5/p_5
1	-1/0.2	0/0.6	1/0.2		
2	-1/0.25	0/0.5	1/0.25		
3	-1/0.1884	0/0.6233	1/0.1884		
4	-1/0.333	0/0.333	1/0.333		

Треба знайти:

1. Вивчити поняття безперервного плану експерименту та інформаційної матриці, а також критерії оптимальності, пов'язані з точністю оцінювання параметрів моделі та точністю оцінювання математичного очікування функції відгуку.

2. Розробити програму з обробки різних планів експерименту для регресійних моделей. Обробка полягає в обчисленні різних характеристик плану, пов'язаних із тим чи іншим критерієм оптимальності.

3. Для кожного з планів обчислити значення функціоналів від інформаційної (дисперсійної) матриць, пов'язаних з такими критеріями, як: D —, A —, E —, Φ_2 —, Λ —, MV —, G -оптимальності. Проранжувати плани, зазначені у варіанті, з позицій різних критеріїв. Вибрати план, найкращий за сукупністю критеріїв. Список планів наведено в табл.2.1.

4. Як спектр плану вибрати один із наведених у табл. 2.1 для відповідної моделі. Ваги точок виразити у вигляді залежності від одного параметра, як у прикладі аналітичної побудови оптимального плану. Для цього параметра визначити допустимі інтервали значень, керуючись тим, що ваги точок мають бути невід'ємними, а кількість таких точок із ненульовими вагами має бути не меншою за меншим за кількість параметрів у моделі. Побудувати графіки зміни критерію оптимальності плану, зазначеного у варіанті, залежно від цього скалярного параметра; визначити за графіком оптимальні значення параметра і критерію. Порівняти отриманий результат із результатами з п. 3.

Розв'язання

Квадратична модель має вигляд:

$$y(x, q) = q_2 x^2 + q_1 x + q_0 \quad -1 \leq x \leq 1$$

Програма:

```
import numpy as np
from scipy.optimize import minimize
import matplotlib.pyplot as plt

# Функція для розрахунку критеріїв оптимальності плану
def optimality_criteria(X, p, reg_coef=1e-9):
    n = X.shape[0]
```

```

k = X.shape[1]

# Обчислення інформаційної матриці
V = np.zeros((k, k))
for i in range(n):
    xi = X[i].reshape(k, 1)
    V += p[i] * np.dot(xi, xi.T)

# Регуляризація інформаційної матриці
V_reg = V + reg_coef * np.identity(k)

# Обчислення всіх критеріїв оптимальності
D = np.linalg.det(V_reg)
A = np.trace(np.linalg.inv(V_reg))
E = np.max(np.linalg.eigvals(V_reg))
F2 = np.trace(np.linalg.inv(V_reg)**2)
L = np.max(np.linalg.eigvals(np.linalg.inv(V_reg)))
MV = np.linalg.det(np.ones((k, k)) + V_reg)
G = np.min(np.linalg.eigvals(V_reg))
return D,A,E,F2,L,MV,G

def experiments():
    # Перший аналіз
    X1 = np.array([[-1, 0, 1],
                   [-1, 0, 1],
                   [-1, 0, 1]])
    p1 = np.array([0.2, 0.6, 0.2])
    experiments_list.append(optimality_criteria(X1,p1))

    # Другий аналіз
    X2 = np.array([[-1, 0, 1],
                   [-1, 0, 1],
                   [-1, 0, 1]])

```

```

p2 = np.array([0.25, 0.5, 0.25])
experiments_list.append(optimality_criteria(X2,p2))

# Третій аналіз
X3 = np.array([[ -1, 0, 1],
               [ -1, 0, 1],
               [ -1, 0, 1]])
p3 = np.array([0.1884, 0.6233, 0.333])
experiments_list.append(optimality_criteria(X3,p3))

# Четвертий аналіз
X4 = np.array([[ -1, 0, 1],
               [ -1, 0, 1],
               [ -1, 0, 1]])
p4 = np.array([0.333, 0.333, 0.333])
experiments_list.append(optimality_criteria(X4,p4))

# Виведення результатів у консоль
def print_results(experiments_list):
    D_list=[]
    A_list=[]
    E_list=[]
    F2_list=[]
    Λ_list=[]
    MV_list=[]
    G_list=[]

    for i in range(len(experiments_list)):
        print("Критерії оптимальності для аналізу #{0}:".format(i+1))
        print("D-оптимальність:", experiments_list[i][0])
        D_list.append(experiments_list[i][0])
        print("A-оптимальність:", experiments_list[i][1])
        A_list.append(experiments_list[i][1])
        print("E-оптимальність:", experiments_list[i][2])
        E_list.append(experiments_list[i][2])

```

```

print("Ф2-оптимальність:", experiments_list[i][3])
F2_list.append(experiments_list[i][3])
print("Λ-оптимальність:", experiments_list[i][4])
Λ_list.append(experiments_list[i][4])
print("MV-оптимальність:", experiments_list[i][5])
MV_list.append(experiments_list[i][5])
print("G-оптимальність:", experiments_list[i][6])
G_list.append(experiments_list[i][6])
print()
return D_list, A_list, E_list, F2_list, Λ_list, MV_list, G_list

# Побудування графіків
def print_plots(experiments_list, D_list, A_list, E_list, F2_list, Λ_list, MV_list, G_list):
    my_list=[D_list, A_list, E_list, F2_list, Λ_list, MV_list, G_list]
    labels = ["D-оптимальність", "A-оптимальність", "E-оптимальність", "Ф2-оптимальність", "Λ-оптимальність", "MV-оптимальність", "G-оптимальність"]
    for i in range(len(my_list)):
        plt.plot(my_list[i])
        plt.title(labels[i])
        plt.xlabel("Значення p")
        plt.ylabel("Значення критерію оптимальності")
        plt.show()

experiments_list=[]
experiments()
D_list, A_list, E_list, F2_list, Λ_list, MV_list, G_list=print_results(experiments_list)
print_plots(experiments_list, D_list, A_list, E_list, F2_list, Λ_list, MV_list, G_list)

```

Результати виконання програми:

Критерії оптимальності для аналізу #1:

D-оптимальність: 2.0000001674807343e-18

А-оптимальність: 1999999917.2596357
Е-оптимальність: 2.000000001
Ф2-оптимальність: 1.499999917259639e+18
Λ-оптимальність: 999999999.9999999
MV-оптимальність: 6.0000004994422165e-09
G-оптимальність: 1e-09

Критерії оптимальності для аналізу #2:

D-оптимальність: 2.0000001674807343e-18
А-оптимальність: 1999999917.2596357
Е-оптимальність: 2.000000001
Ф2-оптимальність: 1.499999917259639e+18
Λ-оптимальність: 999999999.9999999
MV-оптимальність: 6.0000004994422165e-09
G-оптимальність: 1e-09

Критерії оптимальності для аналізу #3:

D-оптимальність: 2.289400191425798e-18
А-оптимальність: 1999999917.2596357
Е-оптимальність: 2.2894000010000006
Ф2-оптимальність: 1.499999917259639e+18
Λ-оптимальність: 999999999.9999999
MV-оптимальність: 6.868201587975257e-09
G-оптимальність: 1e-09

Критерії оптимальності для аналізу #4:

D-оптимальність: 1.9979999454926917e-18
А-оптимальність: 2000000028.281932
Е-оптимальність: 1.9980000010000003

Ф2-оптимальність: $1.5000000282819325e+18$

Λ-оптимальність: 1000000027.7814318

MV-оптимальність: $5.994000942590889e-09$

G-оптимальність: $9.999999717180685e-10$

Побудуємо графіки зміни критерію оптимальності плану, зазначеного у варіанті, та представимо на рис.2.1.1-рис.2.1.7.

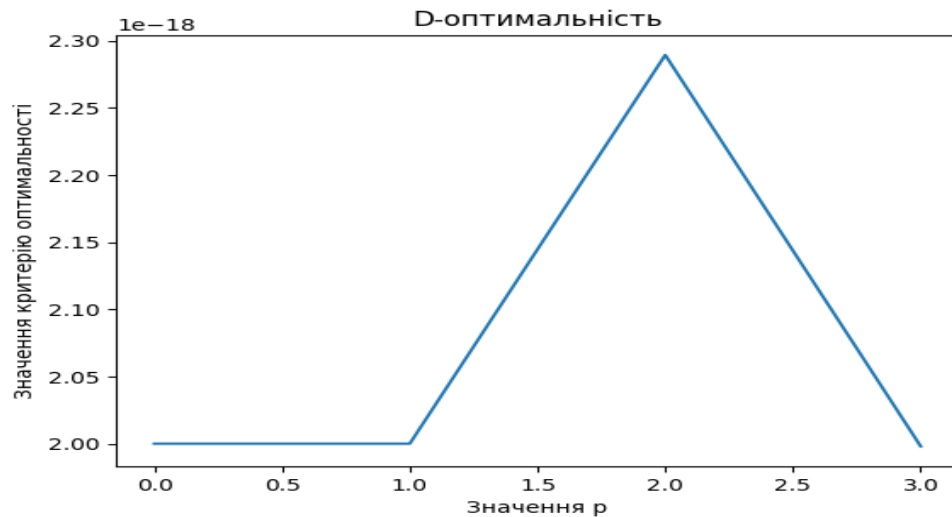


Рисунок 2.1.1 – Залежність критерію оптимальності D від значення p

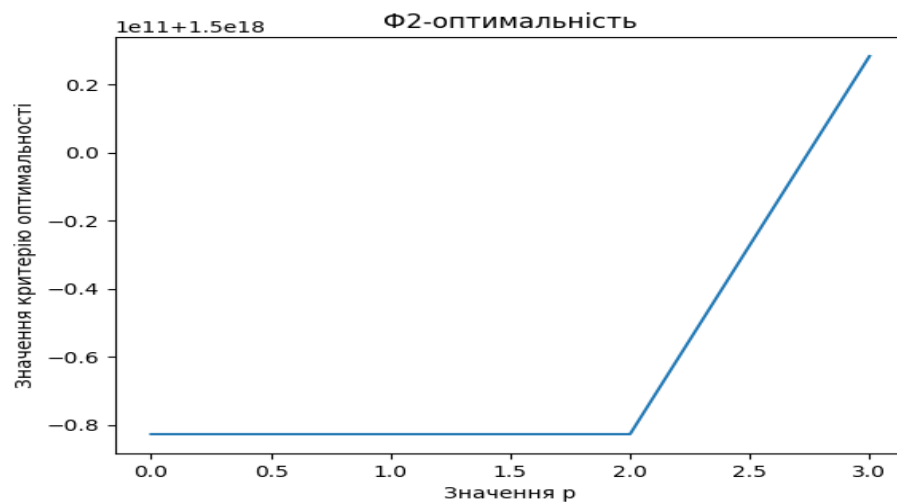


Рисунок 2.1.2 – Залежність критерію оптимальності Ф2 від значення p

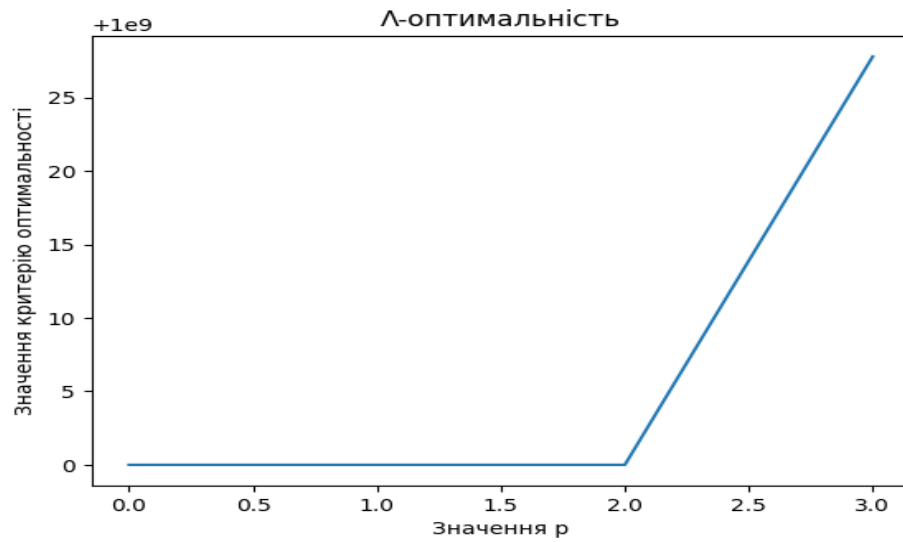


Рисунок 2.1.3 – Залежність критерію оптимальності Λ від значення p

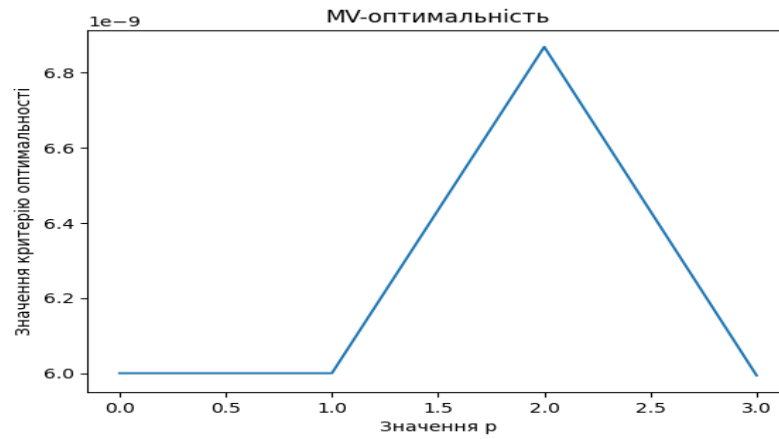


Рисунок 2.1.4 – Залежність критерію оптимальності MV від значення p

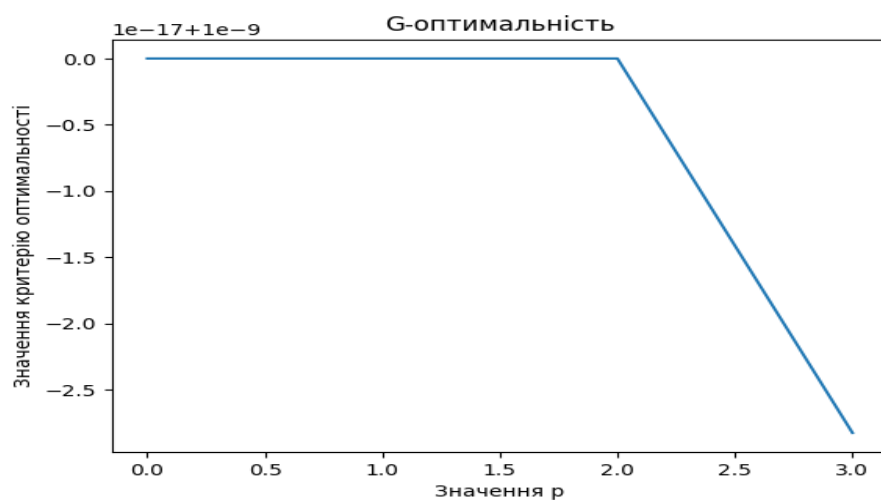


Рисунок 2.1.5 – Залежність критерію оптимальності G від значення p

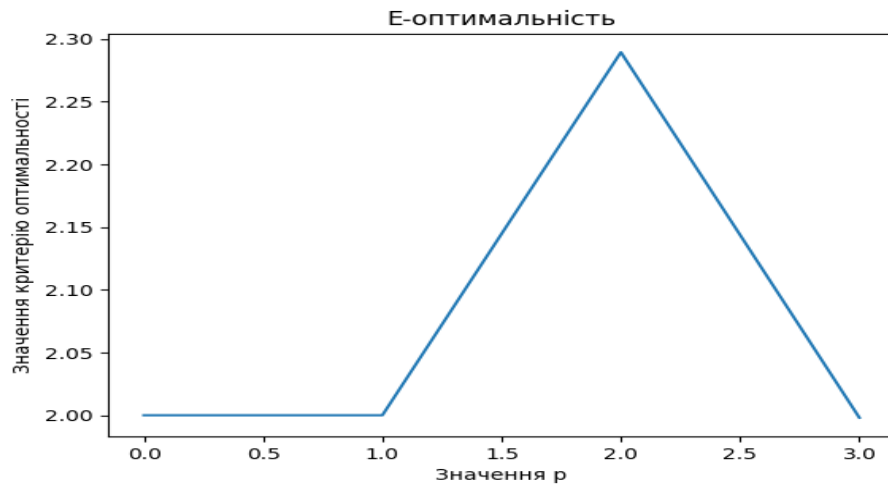


Рисунок 2.1.6 – Залежність критерію оптимальності MV від значення p

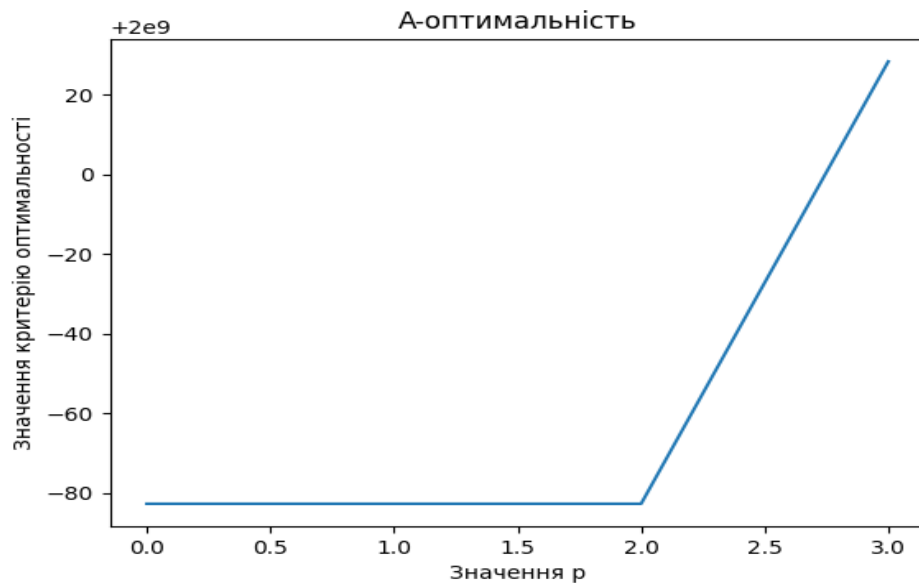


Рисунок 2.1.7 – Залежність критерію оптимальності A від значення p

У процесі виконання практичної роботи я вивчили поняття оптимального плану експерименту та критерії оптимальності планів

Завдання для виконання на практичному занятті

Дана таблиця 2.2., в якій описано плани експерименту.

Таблиця 2.2– Плани експерименту

№ п/п	x_1/p_1	x_2/p_2	x_3/p_3	x_4/p_4	x_5/p_5
1	-1/0.2	0/0.6	1/0.2		
2	-1/0.25	0/0.5	1/0.25		
3	-1/0.1884	0/0.6233	1/0.1884		
4	-1/0.333	0/0.333	1/0.333		
5	-1/0.1273	-0.5/0.3727	0.5/0.3727	1/0.1273	
6	-1 0.152	-0.468/0.348	0.468/0.348	1/0.152	
7	-1/0.1799	-0.5279/0.3201	0.5279/0.3201	1/0.1799	
8	-1/0.25	-0.49/0.25	0.49/0.25	1/0.25	
9	-1/0.093	-0.707/0.248	0/0.3178	0.707/0.248	1/0.093
10	-1/0.107	-0.683/0.25	0/0.286	0.683/0.25	1/0.107
11	-1/0.1092	-0.7379/0.2513	0/0.2785	0.7379/0.2513	1/0.1092
12	-1/0.2	-0.7/0.2	0/0,2	0.7/0.2	1/0.2

Завдання (обираються по варіантам з журналу обліку успішності, або видає викладач)

1. Модель квадратична на відрізку. Плани аналізу: 2-6. Для пункту 4 завдання використовувати критерій D-оптимальності.
2. Модель квадратична на відрізку. Плани аналізу: 5-8. Для пункту 4 завдання використовувати критерій D-оптимальності.
3. Модель квадратична на відрізку, без лінійного члена. Плани аналізу: 1-4. Для пункту 4 завдання використовувати критерій А-оптимальності.
4. Модель квадратична на відрізку, без лінійного члена. Плани аналізу: 5-8. Для пункту 4 завдання використовувати критерій А-оптимальності.

5. Модель кубічна на відрізку. Плани аналізу: 5-8. Для пункту 4 завдання використовувати критерій А-оптимальності.

6. Модель кубічна на відрізку. Плани аналізу: 9-12. Для пункту 4 завдання використовувати критерій А-оптимальності.

7. Модель четвертого ступеня на відрізку. Плани аналізу: 9-12. Для пункту 4 завдання використовувати критерій Φ_2 -оптимальності.

8. Модель четвертого ступеня на відрізку без квадратичного члена. Плани аналізу: 9-12. Для пункту 4 завдання використовувати критерій D-оптимальності.

Для обраного завдання необхідно виконати:

1. Вивчити поняття безперервного плану експерименту та інформаційної матриці, а також критерії оптимальності, пов'язані з точністю оцінювання параметрів моделі та точністю оцінювання математичного очікування функції відгуку.

2. Розробити програму з обробки різних планів експерименту для регресійних моделей. Обробка полягає в обчисленні різних характеристик плану, пов'язаних із тим чи іншим критерієм оптимальності.

3. Для кожного з планів обчислити значення функціоналів від інформаційної (дисперсійної) матриць, пов'язаних з такими критеріями, як: D —, A —, E —, Φ_2 —, Λ —, MV —, G -оптимальності. Проранжувати плани, зазначені у варіанті, з позицій різних критеріїв. Вибрати план, найкращий за сукупністю критеріїв. Список планів наведено в табл.2.1.

4. Як спектр плану вибрати один із наведених у табл. 2.1 для відповідної моделі. Ваги точок виразити у вигляді залежності від одного параметра, як у прикладі аналітичної побудови оптимального плану. Для цього параметра визначити допустимі інтервали значень, керуючись тим, що ваги точок мають бути невід'ємними, а кількість таких точок із ненульовими вагами має бути не меншою за меншим за кількість параметрів у моделі.

Побудувати графіки зміни критерію оптимальності плану, зазначеного у варіанті, залежно від цього скалярного параметра; визначити за графіком оптимальні значення параметра і критерію. Порівняти отриманий результат із результатами з п. 3.

5. Оформити звіт, що містить постановку задачі, отримані результати і текст програми, висновки.

Питання для самостійної роботи

1. Інформаційна матриця та її властивості.
2. Критерії оптимальності планів експерименту, пов'язані з точністю оцінювання параметрів моделі.
3. Критерії оптимальності планів експерименту, пов'язані з точністю оцінювання математичного очікування функції відгуку.
4. Геометрія еліпсоїда розсіювання і критерії оптимальності планів експерименту.

Рекомендована література

[1],[3], [4], [5].

2.2 ПРАКТИЧНА РОБОТА №2 - ПОВНИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ.

Основні теоретичні питання

Мета роботи: По результатам повно факторного експерименту 2^3 для функції відгуку $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{23}x_2x_3 + b_{13}x_1x_3 + b_{123}x_1x_2x_3$ знайти найкращу лінійну незміщену оцінку вектора параметрів $\vec{b} = \{b_0, b_1, \dots, b_m\}$.

Постановка завдання оптимізації.

Пошук оптимальних значень параметрів є однією з важливих завдань, розв'язуваних при створенні нових технічних систем, керуванні виробництвом або технологічними процесами. Відповідно до теорії ефективності необхідно – сформулювати критерій ефективності (функцію відгуку в термінах ТПЕ (техніка планування експерименту)). У більшості випадків ефективність визначається сукупністю показників, що характеризують приватні властивості досліджуваної системи й виконуваної нею операції. Критерій ефективності будується на безлічі значень приватних показників з використанням теорії корисності або методів векторної оптимізації. У деяких випадках критерій ефективності вдається побудувати на безлічі значень одного показника, перевіривши всі інші показники в розряд обмежень; виділити керовані й некеровані параметри (фактори) системи й середовища, що роблять істотний вплив на критерій ефективності; визначити обмеження на значення параметрів.

Завдання оптимізації полягає в знаходженні екстремуму функції відгуку в області припустимих значень параметрів. Щоб знайти екстремум, необхідно мати опис поверхні відгуку в інтервалах варіювання параметрів,

що далеко не завжди вдається одержати виходячи з теоретичних міркувань, тому що функція відгуку в аналітичному виді може бути апіорі невідома.

Реалізація завдання оптимізації, заснована на застосуванні ТПЕ, як і будь-якого завдання експериментального дослідження, починається з визначення об'єкта аналізу, мети дослідження, вивченні сутності досліджуваного процесу, аналізі наявних ресурсів, можливості проведення експериментів з досліджуваним об'єктом у необхідному діапазоні зміни факторів.

Об'єктом аналізу виступає заданий критерій ефективності досліджуваної системи, розглянутий як функція від істотних параметрів системи й зовнішнього середовища. Система може являти собою реальний фізичний об'єкт або його модель – фізичну або математичну (імітаційну, складну аналітичну).

Вивчення процесу функціонування об'єкта дозволяє виявити фактори, що роблять істотний вплив на функцію відгуку. Вибір істотних змінних потенційно визначає ступінь досягнення адекватності одержуваної моделі: відсутність у вихідному переліку істотних параметрів, так ще й доволіно мінливих у ході експерименту, не дозволяє правильно вирішити завдання оптимізації; включення несуттєвих параметрів ускладнює модель, викликає значне збільшення обсягу експериментів, хоча за результатами дослідження неістотність відповідних параметрів буде виявлена.

Для кожної змінної варто визначити діапазон і характер зміни (безперервність або дискретність). Обмеження на діапазон змін можуть носити принциповий або технічний характер. Принципові обмеження факторів не можуть бути порушені при будь-яких обставинах. Ці обмеження задаються виходячи з фізичних подань (наприклад, ємність пристроїв пам'яті завжди має позитивне значення). Другий тип обмежень пов'язаний з техніко-економічними міркуваннями, наприклад, з наявністю відповідного

апаратно-програмного комплексу, прийнятою технологією обробки інформації.

Виділення області зміни факторів є не формальним завданням, а ґрунтується на досвіді дослідника. У рамках області припустимих значень факторів необхідно виділити початкову область планування експерименту. Цей вибір включає визначення основного (нульового) рівня як вихідної точки побудови плану й інтервалів варіювання. Інтервал варіювання задає щодо основного рівня значення фактора, при яких будуть вироблятися експерименти. Звичайно інтервали є симетричними щодо центрального значення. Інтервал варіювання повинен відповідати двом обмеженням: його застосування не повинне приводити до виходу фактора за межі області припустимих значень; він повинен бути більше погрішності завдання значень фактора (у протилежному випадку рівні фактора стануть не помітними). У межах цих обмежень вибір конкретного значення є неформальною процедурою, що враховує орієнтовну інформацію про кривизну поверхні функції відгуку.

Фактор повинен бути керованим, тобто експериментатор може підтримувати його постійне значення протягом усього досвіду. Для фактора необхідно вказати його конкретні значення й засоби контролю. Сам фактор повинен бути первинним, тому що складно управляти фактором, що у свою чергу є функцією інших факторів. Для кожного фактора варто вказати точність його завдання й підтримки в ході експерименту.

Одночасна зміна факторів припускає їхню сумісність, що означає реалізація і безпека всіх їхніх сполучень. Необхідно також забезпечити незалежність зміни кожного фактора, що означає можливість установалення будь-якого значення фактора поза зв'язком зі значеннями інших факторів.

Ціль дослідження, необхідна точність одержуваних результатів, наявні ресурси обмежують безліч припустимих моделей функції відгуку (з ускладненням моделі й підвищенням точності оцінки показників різко

зростає обсяг необхідних досвідів) і відповідно визначають план проведення експериментів.

Повний факторний експеримент типу 2^k .

На початкових етапах оптимізації для визначення градієнта застосовують неповні поліноми другого порядку або лінійні поліноми. Обчислення оцінок коефіцієнтів таких поліномів здійснюється на основі обробки результатів реалізації найбільш простих планів, у яких кожен фактор приймає тільки два значення $v_{i, \min}$ або $v_{i, \max}$, розташовані симетрично щодо деякого нульового рівня або центра плану по даному факторі. Значення рівнів варіювання вибирає дослідник, виходячи з можливого діапазону зміни кожного фактора й можливості застосування лінійної апроксимації функції відгуку в обраному діапазоні змін параметра. Без обмеження спільності можна вважати, що кодовані значення x_i приймають значення -1 й $+1$ відповідно (або просто $-$ або $+$). Безліч всіх крапок в k -мірному просторі, координати яких є комбінаціями "+" й "-", називається повним факторним планом або планом *повного факторного експерименту* типу 2^k (ПФЕ). Кількість крапок у цьому плані $N = 2^k$.

Приклади розв'язання

Приклад 1. Візьмемо повний факторний експеримент із трьома незалежними змінними x_1 , x_2 і x_3 , табл. 2.2.1. Другий, третій і четвертий стовпці таблиці відповідають властивостям плану експериментів, п'ятий – восьмий стовпці містять значення добутків незалежних змінних. Фіктивна змінна $x_0 = 1$ (перший стовпець) уведена для однаковості запису розрахункових формул коефіцієнтів полінома.

Таблиця 2.2.1 - Повний факторний експеримент із трьома незалежними змінними x_1 , x_2 і x_3

Матриця планування								Вектор результатів
1	2	3	4	5	6	7	8	9
x_0	x_1	x_2	x_3	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	$x_1 x_2 x_3$	y
+	—	—	—	+	+	+	—	y_1
+	—	—	+	+	—	—	+	y_2
+	—	+	—	—	+	—	+	y_3
+	—	+	+	—	—	+	—	y_4
+	+	—	—	—	—	+	+	y_5
+	+	—	+	—	+	—	—	y_6
+	+	+	—	+	—	—	—	y_7
+	+	+	+	+	+	+	+	y_8

Рядки відповідають дослідженням, наприклад, перший рядок характеризує експеримент, у якому всі незалежні змінні перебувають на нижньому рівні.

Існує кілька способів побудови подібних матриць планування. Зокрема можна скористатися прийомом, характерним для запису послідовності двійкових чисел.

У стовпці останньої змінної x_3 знаки міняються по черзі, у стовпці передостанньої змінної x_2 — чергуються через два елементи, третьої праворуч змінної x_1 — через чотири елементи.

Аналогічно будується матриця для будь-якої кількості змінних, порядок перерахування змінних не грає ролі. Стовпці з добутками змінних обчислюються шляхом множення значень елементів у відповідних стовпцях простих змінних.

З аналізу матриці планування легко бачити, що повний факторний експеримент має властивості:

Ортогональності. Сума парних добутків елементів будь-яких двох різних стовпців дорівнює нулю. Зокрема, для простих змінних:

$$\sum_{u=1}^N x_{iu} x_{ju} = 0, \quad i \neq j, \quad i, j = \overline{0, k}.$$

Симетричності. Сума всіх елементів будь-якого стовпця, за винятком першого, дорівнює нулю, наприклад:

$$\sum_{u=1}^N x_{iu} = 0, \quad i = \overline{1, k}.$$

Нормованості. Сума квадратів елементів будь-якого стовпця дорівнює числу досвідів, так для i -ї змінної:

$$\sum_{u=1}^N x_{iu}^2 = N, \quad i = \overline{0, k}.$$

Перші дві властивості забезпечують незалежність оцінок коефіцієнтів моделі й допустимість їхньої фізичної інтерпретації. Порухення цих властивостей приводить до взаємної залежності оцінок і неможливості додання змісту коефіцієнтам.

Включення в матрицю планування змінних виду x_i^2 приведе до появи одиничних стовпців, що збігаються один з одним і зі стовпцем x_0 . Отже, не можна буде визначити, за рахунок чого отримане значення β_0 . Тому плани ПФЕ 2^k не застосовні для побудови функції відгуку у вигляді повного полінома другого ступеня.

Оцінки коефіцієнтів функції відгуку.

За допомогою матриці планування, описаної в таблиці 6.1, можна обчислити оцінки коефіцієнтів неповного полінома третього ступеня:

$$y' = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{123} x_1 x_2 x_3,$$

або лінійної функції:

$$y' = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3.$$

Перший вид полінома дозволяє оцінити не тільки вплив окремих факторів, але й один із часто, що зустрічаються видів, нелінійності, коли

ефект одного фактора залежить від рівня інших факторів, тобто є присутнім ефект взаємодії факторів. Ефект взаємодії виду $x_i x_j$ називають парним, $x_i x_j x_k$ – потрійним і т.д. З ростом кількості факторів число можливих взаємодій швидко збільшується. Сумарно кількість всіх коефіцієнтів функції відгуку такого типу дорівнює числу дослідів повного факторного експерименту.

Оцінки коефіцієнтів полінома визначаються на основі методу найменших квадратів і для розглянутого типу ПФЕ обчислюються по простих співвідношеннях:

$$\beta_i = \frac{1}{N} \sum_{u=1}^N x_{iu} \bar{y}_u, \quad i = \overline{0, k}$$

$$\beta_{i, \dots, m} = \frac{1}{N} \sum_{u=1}^N x_{iu} \dots x_{mu} \bar{y}_u, \quad i = \overline{1, k}, \quad m > i,$$

Тут величина y відповідає значенню відгуку в зазначеній крапці факторного простору при відсутності повторних дослідів або є оцінкою математичного очікування:

$$\bar{y}_u = \frac{1}{r_u} \sum_{i=1}^{r_u} y_{ui}$$

значення функції відгуку по всім r_u повторним дослідженням у даній крапці.

Повторні дослідження проводяться в тих випадках, коли на функціонування системи впливають випадкові впливи. Кількість повторних досліджень у різних крапках плану може розрізнятися.

Припустимо наступну інтерпретацію оцінок коефіцієнтів:

b_0 - відповідає значенню функції відгуку в центрі проведеного експерименту;

b_i - дорівнює збільшенню функції при переході значення фактора i з нульового рівня на верхній (це внесок відповідного фактора в значення функції);

b_{ij} - дорівнює нелінійної частини збільшення функції при одночасному переході факторів i й j з нульового рівня на верхній і т.п.

Помилки у визначенні коефіцієнтів полінома можна охарактеризувати відповідною дисперсією. З обліком того, що кодовані значення факторів приймають значення $+1$ й -1 , оцінка дисперсії коефіцієнта визначається співвідношенням:

$$\sigma^2(\beta_i) = D\left[\frac{1}{N} \sum_{u=1}^N x_u \bar{y}_u\right] = \frac{1}{N^2} \sum_{u=1}^N D(\bar{y}_u) = \frac{1}{N} \left[\frac{1}{N} \sum_{u=1}^N D(\bar{y}_u) \right],$$

Отже, оцінка дисперсії всіх коефіцієнтів однакова й визначається тільки дисперсією середніх значень функції відгуку й числом досвідів. Цю формулу можна застосовувати, якщо кількість досвідів у всіх крапках плану однакова. При факторному експерименті, на відміну від класичного, одночасно варіюються всі фактори, тому кожен коефіцієнт полінома визначається за результатами всіх експериментів, тим самим оцінка дисперсії коефіцієнтів виходить в N раз менше середньої дисперсії всіх досвідів. Оцінка дисперсії середнього значення в конкретній крапці плану:

$$D(\bar{y}_u) = \frac{\sigma_u^2}{r_u}$$

де σ_u^2 – оцінка дисперсії функції відгуку в крапці u ,
 r_u – число повторних досвідів у цій крапці плану.

Дисперсія оцінок всіх коефіцієнтів однакова, тому ПФЕ розглянутого типу є ротатабельним.

При використанні неповних поліномів k -го порядку кількість крапок плану дорівнює кількості оцінюваних параметрів (насичене планування). Тому не залишається ступенів волі для перевірки гіпотези про адекватність подання результатів експерименту заданою математичною моделлю. Якщо

застосовувати поліноми першого ступеня, то тоді залишаються ступені волі для перевірки гіпотези про адекватність моделі.

Приклад 2.

За результатами повнофакторного експерименту (табл.2.2.2) По результатам повно факторного експерименту 2^3 для функції відгуку $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{23}x_2x_3 + b_{13}x_1x_3 + b_{123}x_1x_2x_3$ знайти найкращу лінійну незміщену оцінку вектора параметрів $\vec{b} = \{b_0, b_1, \dots, b_m\}$.

Таблиця 2.2.2- Результатам повно факторного експерименту 2^3 для функції відгуку

№ крапок плану	Фактори								Параметри оптимізації		
	$b_0,$ j=0	$x_1,$ j=1	$x_2,$ j=2	$x_3,$ j=3	$x_1x_2,$ j=4	$x_2x_3,$ j=5	$x_3x_1,$ j=6	$x_1x_2x_3,$ j=7	$y'_i,$ p=1	$y''_i,$ p=2	$\bar{y}_i$
1	+	+	+	+	+	+	+	+	58	50	
2	+	-	+	+	-	+	-	-	6	12	
3	+	+	-	+	-	-	+	-	8	9	
4	+	-	-	+	+	-	-	+	0	2	
5	+	+	+	-	+	-	-	-	14	15	
6	+	-	+	-	-	-	+	+	0	-2	
7	+	+	-	-	-	+	-	+	2	3	
8	+	-	-	-	+	+	+	-	-3	-4	

Визначити :

1. Знайти середній відгук

$$\bar{y}_i = \frac{\sum_{p=1}^r \bar{y}_{ip}}{r}.$$

2. Знайти найбільш імовірні оцінки (найкращу лінійну незміщену оцінку) коефіцієнтів моделі

$$b_j = \frac{\sum_{i=1}^N x_{ij} \bar{y}_i}{N}.$$

3. Показати, що ця розрахункова формула коефіцієнтів регресії є наслідком застосування методу максимальної правдоподібності у випадку, коли інформаційна матриця виходить із матриці плану повно факторного експерименту, стовпці якої попарно ортогональні.

4. Знайти порядкові дисперсії:

$$(\sigma_i)^2 = \frac{\sum_{p=1}^r (y_{ip} - \bar{y}_i)^2}{r-1}$$

й дисперсію відтворюваності:

$$S^2(y) = \frac{1}{N} \sum_{i=1}^N (\sigma_i)^2 = \frac{\sum_{i=1}^N \sum_{p=1}^r (y_{ip} - \bar{y}_i)^2}{N(r-1)}.$$

5. По Т-критерії Стюдента перевірити значимість коефіцієнтів регресії. Для цього знайти дисперсію коефіцієнтів регресії:

$$S^2(b_j) = \frac{S^2(y)}{Nr},$$

по таблицях визначити критичне значення $T_{кр}$ критерію Стюдента для ступенів волі $f = N \cdot (r-1)$ і рівня значимості $\alpha = 5\%$.

Побудувати інтервал довжиною $r\Delta b_j$, $\Delta b_j = T_{кр} S(b_j)$, де $\Delta b_j = T_{кр} S(b_j)$ - половина довжини довірчого інтервалу. Якщо абсолютна величина коефіцієнта $|b_j| > \Delta b_j$, то коефіцієнт вважається значущим.

6. Перевірити *адекватність* моделі на основі критерію Фішера. Із цією метою обчислити дисперсію адекватності

$$S^2_{ad} = \frac{r}{N-m} \sum_{i=1}^N (\bar{y}_i - \bar{y})^2,$$

де m – кількість членів апроксимуючого полінома, з огляду на й вільний член;

$\hat{y}_i$ – значення функції відгуку, обчислене по прийнятій функції регресії.

Обчислити значення критерію Фішера:

$$F = \frac{S^2_{ad}}{S^2(y)}.$$

З таблиць знайти критичне значення критерію Фішера $F_{кр}$ для ступенів волі $F_{ad}=N-m$, $f_k=N(r-1)$ і заданого рівня значимості $\alpha=5\%$.

При виконанні умови $F < F_{кр}$ модель зізнається адекватної. б) Обчислити коефіцієнт множинної кореляції:

$$R = \sqrt{1 - \frac{\sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2}{\sum_{i=1}^N (\bar{y}_i - \bar{y})^2}}.$$

Якщо R близько до одиниці, то рівняння регресії добре описує експериментальні дані.

7. Виконати звіт на тему завдання згідно варіанта, умову, а також відповідні розрахунки та лістинг програмного коду виконання завдання. Повині бути описані дії та хід виконання роботи. Після закінчення виконання завдання, по отриманим даним зробити висновки (детальний опис див. в завданні).

Розв'язання:

Для цього скористаємось мовою програмування Python, спочатку задамо початкові дані.

```
import numpy as np
import statsmodels.api as sm
```

```

x = np.array([[1, 1, 1, 1, 1, 1, 1, 1],
[1, -1, 1, -1, 1, -1, 1, 1],
[1, 1, -1, -1, -1, 1, -1, -1],
[1, -1, -1, 1, -1, -1, -1, 1],
[1, 1, 1, -1, -1, -1, -1, 1],
[1, 1, -1, 1, -1, -1, 1, -1],
[1, 1, -1, -1, -1, 1, -1, -1],
[1, 1, -1, -1, -1, 1, -1, -1]])
y1=np.array([58, 6, 8, 0, 14, 0, 2, -3])
y2=np.array([50, 12, 9, 2, 15, -2, 3, -4])

```

Тепер знайдемо середній відгук.

```

average_response1 = np.mean(y1)
average_response2 = np.mean(y2)
print(average_response1) # 10.625
print(average_response2) # 10.625

```

Далі скористаємося модулем `statmodels.OLS` – методом найменших квадратів. Метод найменших квадратів (МНК) – це статистичний метод для оцінки параметрів моделі лінійної регресії. Він використовується для знаходження оптимальних значень коефіцієнтів моделі, мінімізуючи суму квадратів відхилень передбачених значень від фактичних. У моделі лінійної регресії припускається, що залежна змінна є лінійною комбінацією незалежних змінних із деякими коефіцієнтами. Мета МНК - знайти найкращі значення цих коефіцієнтів, щоб мінімізувати суму квадратів помилок, які виникають під час передбачення залежної змінної на основі незалежних змінних.

```

# Додаємо константу до модуля sm
X = sm.add_constant(x)
print(X)

```

Додавання константи до модуля sm представимо на рис.2.2.1

```
[[ 1  1  1  1  1  1  1  1]
 [ 1 -1  1 -1  1 -1  1  1]
 [ 1  1 -1 -1 -1  1 -1 -1]
 [ 1 -1 -1  1 -1 -1 -1  1]
 [ 1  1  1 -1 -1 -1 -1  1]
 [ 1  1 -1  1 -1 -1  1 -1]
 [ 1  1 -1 -1 -1  1 -1 -1]
 [ 1  1 -1 -1 -1  1 -1 -1]]
```

Рисунок 2.2.1 – Додавання константи до модуля sm

Далі побудуємо лінійну регресійну модель та отримаємо найімовірніші оцінки коефіцієнтів моделі.

```
model1 = sm.OLS(y1, X).fit()
```

```
model2 = sm.OLS(y2, X).fit()
```

```
print("Параметри першої моделі", model1.params) # [11.57608696  8.8804347
8  5.52898551  7.40942029  3.74637681  9.71014493  1.13405797 10.01449275]
```

```
print("Параметри другої моделі", model1.params) # [11.57608696  8.88043478
5.52898551  7.40942029  3.74637681  9.71014493  1.13405797 10.01449275]
```

Далі знайдемо порядкові дисперсії та дисперсії відтворюваності, порядкові дисперсії можна знайти як суму квадратів відхилень кожного відгуку від середнього відгуку, поділену на кількість повторень кожної комбінації факторів. Діагональні елементи коваріаційної матриці дають дисперсії коефіцієнтів регресії. Ми можемо отримати їх наступним чином:

```
# Порядкові дисперсії
```

```
ordinal_variances1 = model1.mse_resid * np.diag(np.linalg.inv(np.dot(x.T, x)))
```

```
ordinal_variances2 = model2.mse_resid * np.diag(np.linalg.inv(np.dot(x.T, x)))
```

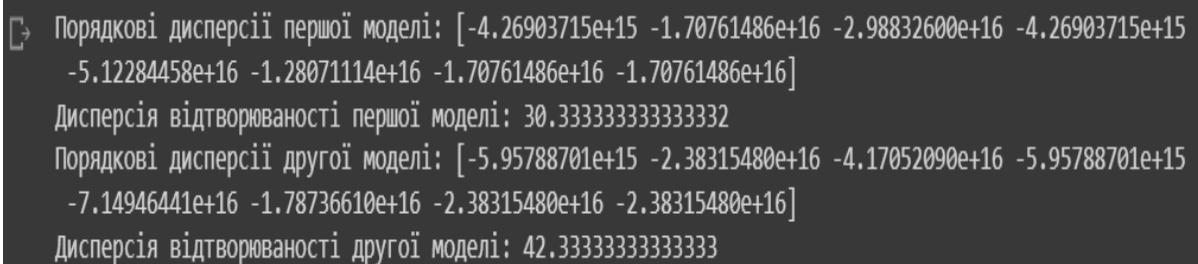
```
# Дисперсія відтворюваності
```

```
reproducibility_variance1 = model1.mse_resid
```

```
reproducibility_variance2 = model2.mse_resid
```

```
print("Порядкові дисперсії першої моделі:", ordinal_variances1)  
print("Дисперсія відтворюваності першої моделі:", reproducibility_variance1)  
print("Порядкові дисперсії другої моделі:", ordinal_variances2)  
print("Дисперсія відтворюваності другої моделі:", reproducibility_variance2)
```

Результат розрахунку, представимо на рис.2.2.2-2.2.6.



```
Порядкові дисперсії першої моделі: [-4.26903715e+15 -1.70761486e+16 -2.98832600e+16 -4.26903715e+15  
-5.12284458e+16 -1.28071114e+16 -1.70761486e+16 -1.70761486e+16]  
Дисперсія відтворюваності першої моделі: 30.33333333333332  
Порядкові дисперсії другої моделі: [-5.95788701e+15 -2.38315480e+16 -4.17052090e+16 -5.95788701e+15  
-7.14946441e+16 -1.78736610e+16 -2.38315480e+16 -2.38315480e+16]  
Дисперсія відтворюваності другої моделі: 42.33333333333333
```

Рисунок 2.2.2 – Результат розрахунку порядкових дисперсій та дисперсії відтворюваності

Для отримання дисперсії коефіцієнтів регресії спочатку отримаємо коваріаційну матрицю коефіцієнтів регресії. Діагональні елементи коваріаційної матриці дають дисперсії коефіцієнтів регресії.

```
cov_matrix1 = model1.cov_params()  
variances1 = cov_matrix1.diagonal()  
cov_matrix2 = model2.cov_params()  
variances2 = cov_matrix2.diagonal()  
print("Дисперсії коефіцієнтів регресії першої моделі", variances1)  
print("Дисперсії коефіцієнтів регресії другої моделі", variances2)
```

```
Дисперсії коефіцієнтів регресії першої моделі [4.92414934 4.73779143 2.67829763 4.15482567 1.73934047 4.22650179
3.45239971 2.77147658]
Дисперсії коефіцієнтів регресії другої моделі [6.87216446 6.61208255 3.73784394 5.79849296 2.42743121 5.89852447
4.8181842 3.8678849 ]
```

Рисунок 2.2.3 – Результат розрахунку дисперсій коефіцієнтів регресії

Отримаємо t-статистику та p-значення для кожного коефіцієнта та перевіримо нульову гіпотезу про те, що кожен коефіцієнт регресії дорівнює нулю.

```
# Отримаємо t-статистику та p-значення для кожного коефіцієнта
```

```
tvals1 = model1.tvalues
```

```
pvals1 = model1.pvalues
```

```
print("t-статистика першого модуля", tvals1)
```

```
print("p-значення першого модуля", pvals1)
```

```
tvals2 = model2.tvalues
```

```
pvals2 = model2.pvalues
```

```
print("t-статистика другого модуля", tvals1)
```

```
print("p-значення другого модуля", pvals1)
```

```
# Перевіримо нульову гіпотезу про те, що кожен коефіцієнт регресії дорівнює нулю
```

```
print("Перший набір даних")
```

```
for i in range(len(model1.params)):
```

```
    if pvals1[i] < 0.05:
```

```
        print("Коефіцієнт {} є статистично значущим на рівні 0,05 з t-значенням = {} та p-значенням = {}".format(i, tvals1[i], pvals1[i]))
```

```
    else:
```

```
        print("Коефіцієнт {} не є статистично значущим на рівні 0,05 з t-значенням = {} та p-значенням = {}".format(i, tvals1[i], pvals1[i]))
```

```
print("Другий набір даних")
```

```

for i in range(len(model2.params)):
    if pvals2[i] < 0.05:
        print("Коефіцієнт {} є статистично значущим на рівні 0,05 з t-
значенням = {} та p-значенням = {}".format(i, tvals2[i], pvals2[i]))
    else:
        print("Коефіцієнт {} не є статистично значущим на рівні 0,05 з t-
значенням = {} та p-значенням = {}".format(i, tvals2[i], pvals2[i]))

```

```

t-статистика першого модуля [5.21670373 4.07986939 3.37843862 3.63502856 2.84065934 4.72318747
0.61034406 6.01552242]
p-значення першого модуля [0.03483697 0.05515368 0.0775574 0.06804655 0.10480317 0.04202075
0.60374953 0.02653943]
t-статистика другого модуля [5.21670373 4.07986939 3.37843862 3.63502856 2.84065934 4.72318747
0.61034406 6.01552242]
p-значення другого модуля [0.03483697 0.05515368 0.0775574 0.06804655 0.10480317 0.04202075
0.60374953 0.02653943]
Перший набір даних
Коефіцієнт 0 є статистично значущим на рівні 0,05 з t-значенням = 5.216703728658587 та p-значенням = 0.034836973570372895
Коефіцієнт 1 не є статистично значущим на рівні 0,05 з t-значенням = 4.079869390330677 та p-значенням = 0.055153684981440576
Коефіцієнт 2 не є статистично значущим на рівні 0,05 з t-значенням = 3.3784386225366227 та p-значенням = 0.07755740131882803
Коефіцієнт 3 не є статистично значущим на рівні 0,05 з t-значенням = 3.6350285594925196 та p-значенням = 0.06804655178473222
Коефіцієнт 4 не є статистично значущим на рівні 0,05 з t-значенням = 2.8406593406593337 та p-значенням = 0.10480317335571611
Коефіцієнт 5 є статистично значущим на рівні 0,05 з t-значенням = 4.723187468907218 та p-значенням = 0.04202074532831108
Коефіцієнт 6 не є статистично значущим на рівні 0,05 з t-значенням = 0.6103440568036816 та p-значенням = 0.603749529369751
Коефіцієнт 7 є статистично значущим на рівні 0,05 з t-значенням = 6.0155224246679095 та p-значенням = 0.026539428889546007
Другий набір даних
Коефіцієнт 0 є статистично значущим на рівні 0,05 з t-значенням = 4.5651254816570015 та p-значенням = 0.0447849735282598
Коефіцієнт 1 не є статистично значущим на рівні 0,05 з t-значенням = 2.4644022354364594 та p-значенням = 0.13266548308683948
Коефіцієнт 2 не є статистично значущим на рівні 0,05 з t-значенням = 2.4812374101650767 та p-значенням = 0.13120890725347428
Коефіцієнт 3 не є статистично значущим на рівні 0,05 з t-значенням = 1.924437872659487 та p-значенням = 0.19418628974635468
Коефіцієнт 4 не є статистично значущим на рівні 0,05 з t-значенням = 2.4231771831951776 та p-значенням = 0.13632870487088536
Коефіцієнт 5 не є статистично значущим на рівні 0,05 з t-значенням = 3.305894434094051 та p-значенням = 0.08059353190114549
Коефіцієнт 6 не є статистично значущим на рівні 0,05 з t-значенням = 0.48363401543243995 та p-значенням = 0.6764176732295245
Коефіцієнт 7 є статистично значущим на рівні 0,05 з t-значенням = 4.77885792110819 та p-значенням = 0.041106605054671924

```

Рисунок 2.2.4 – Результат розрахунку t-статистики та p-значень

Тепер перевіримо адекватність моделі на основі критерію Фішера. Критерій Фішера – це статистичний тест, який використовується для оцінки загальної адекватності лінійної регресійної моделі. Він базується на F-статистиці, яка порівнює варіацію залежної змінної, що пояснюється

регресійною моделлю, з варіацією, яка не пояснюється моделлю. Для перевірки адекватності моделі на основі критерію Фішера потрібно розрахувати F-статистику і порівняти її з критичним значенням з F-розподілу.

Нульовою гіпотезою для цього тесту є те, що модель не є значущою, тобто жодна з незалежних змінних не має значущого впливу на залежну змінну.

```
from scipy.stats import f
# Обчислимо F-статистику, використовуючи атрибут fvalue моделі
f_statistic1 = model1.fvalue
f_statistic2 = model2.fvalue
print("F-статистика першої моделі", f_statistic1)
print("F-статистика другої моделі", f_statistic2)

# Обчислимо ступені свободи для чисельника та знаменника F-
статистики, використовуючи атрибути моделі df_model та df_resid відповід
но
dfn1 = model1.df_model
dfd1 = model1.df_resid
dfn2 = model2.df_model
dfd2 = model2.df_resid

# Обчислимо критичне значення F-
розподілу за допомогою методу ppf з модуля scipy.stats
alpha = 0.05
f_critical1 = f.ppf(1 - alpha, dfn1, dfd1)
f_critical2 = f.ppf(1 - alpha, dfn2, dfd2)
print("Критичне значення F-розподілу першої моделі", f_critical1)
print("Критичне значення F-розподілу другої моделі", f_critical2)
```

```

# Порівняємо розраховану F-статистику з критичним значенням F-
розподілу.
# Якщо розрахована F-
статистика більша за критичне значення, ми відкидаємо нульову гіпотезу і
робимо висновок, що модель є значущою.
print("Перший набір даних")
if f_statistic1 > f_critical1:
    print("Модель є значущою за критерієм Фішера.")
else:
    print("Модель не є значущою за критерієм Фішера.")

print("Другий набір даних")
if f_statistic2 > f_critical2:
    print("Модель є значущою за критерієм Фішера.")
else:
    print("Модель не є значущою за критерієм Фішера.")

```

```

F-статистика першої моделі 17.86291208791209
F-статистика другої моделі 9.426181102362206
Критичне значення F-розподілу першої моделі 19.296409652017232
Критичне значення F-розподілу другої моделі 19.296409652017232
Перший набір даних
Модель не є значущою за критерієм Фішера.
Другий набір даних
Модель не є значущою за критерієм Фішера.

```

Рисунок 2.2.5 – Результат перевірки адекватності моделі на основі критерію Фішера

Обчислимо коефіцієнт множинної кореляції, якщо він близько до одиниці, то рівняння регресії добре описує експериментальні дані.

```

# Отримаємо значення R-
квадрат (також квадрат коефіцієнта множинної кореляції)
R_squared1 = model1.rsquared
R_squared2 = model2.rsquared

# Візьмемо квадратний корінь, щоб отримати коефіцієнт множинної кореляції
multiple_R1 = R_squared1 ** 0.5
multiple_R2 = R_squared2 ** 0.5
print("Коефіцієнт множинної кореляції першої моделі", multiple_R1)
print("Коефіцієнт множинної кореляції другої моделі", multiple_R2)

```

Коефіцієнт множинної кореляції першої моделі 0.9889882142483492
Коефіцієнт множинної кореляції другої моделі 0.9794347479164742

Рисунок 2.2.6 – Результат розрахунку коефіцієнту множинної кореляції

Висновок: У процесі виконання практичної роботи я виконав повно факторний експеримент 2^3 для функції відгуку $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{23}x_2x_3 + b_{13}x_1x_3 + b_{123}x_1x_2x_3$ та знайшов найкращу лінійну незміщену оцінку вектора параметрів $\vec{b} = \{b_0, b_1, \dots, b_m\}$. Побудував лінійну регресійну модель та отримав найімовірніші оцінки коефіцієнтів моделі. Знайшов середній відгук, порядкові дисперсії та дисперсії відтворюваності. Отримав t-статистику та p-значення для кожного коефіцієнта та перевірів нульову гіпотезу про те, що кожен коефіцієнт регресії дорівнює нулю. Після цього перевірів адекватність моделі на основі критерію Фішера та розрахував коефіцієнт множинної кореляції який виявився близьким до 1, отже рівняння регресії добре описує експериментальні дані.

Завдання для виконання на практичному занятті

В ході виконання практичного завдання треба проаналізувати поняття планування.

За результатами повно факторного експерименту 2^n для функції відгуку $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{23}x_2x_3 + b_{13}x_1x_3 + b_{123}x_1x_2x_3$ знайти найкращу *лінійну незміщену* оцінку вектора параметрів $\hat{b} = \{b_0, b_1, \dots, b_m\}$.

Для цього потрібно:

1. Знайти середній відгук

$$\bar{y}_i = \frac{\sum_{p=1}^r \bar{y}_{ip}}{r}.$$

2. Знайти найбільш імовірні оцінки (*найкращу лінійну незміщену оцінку*) коефіцієнтів моделі

$$b_j = \frac{\sum_{i=1}^N x_{ij} \bar{y}_i}{N}.$$

3. Показати, що ця розрахункова формула коефіцієнтів регресії є наслідком застосування методу максимальної правдоподібності у випадку, коли інформаційна матриця виходить із матриці плану повно факторного експерименту, стовпці якої попарно ортогональні.

4. Знайти порядкові дисперсії:

$$(\sigma_i)^2 = \frac{\sum_{p=1}^r (y_{ip} - \bar{y}_i)^2}{r-1}$$

й дисперсію відтворюваності:

$$S^2(y) = \frac{1}{N} \sum_{i=1}^N (\sigma_i)^2 = \frac{\sum_{i=1}^N \sum_{p=1}^r (y_{ip} - \bar{y}_i)^2}{N(r-1)}.$$

5. По Т-критерію Стюдента перевірити значимість коефіцієнтів регресії. Для цього знайти дисперсію коефіцієнтів регресії:

$$S^2(b_j) = \frac{S^2(y)}{Nr},$$

по таблицях визначити критичне значення $T_{кр}$ критерію Стюдента для ступенів волі $f = N \cdot (r-1)$ і рівня значимості $\alpha = 5\%$.

Побудувати інтервал довжиною $r\Delta b_j$, $\Delta b_j = T_{кр} S(b_j)$, де $\Delta b_j = T_{кр} S(b_j)$ - половина довжини довірчого інтервалу. Якщо абсолютна величина коефіцієнта $|b_j| > \Delta b_j$, то коефіцієнт вважається значущим.

6. Перевірити *адекватність* моделі на основі критерію Фішера. Із цією метою обчислити дисперсію адекватності

$$S^2_{ad} = \frac{r}{N-m} \sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2,$$

де m – кількість членів апроксимуючого полінома, з огляду на й вільний член;

$\hat{y}_i$ – значення функції відгуку, обчислене по прийнятій функції регресії.

Обчислити значення критерію Фішера:

$$F = \frac{S^2_{ad}}{S^2(y)}.$$

З таблиць знайти критичне значення критерію Фішера $F_{кр}$ для ступенів волі $F_{ad} = N-m$, $f_k = N(r-1)$ і заданого рівня значимості $\alpha = 5\%$.

При виконанні умови $F < F_{кр}$ модель зізнається адекватною. б) Обчислити коефіцієнт множинної кореляції:

$$R = \sqrt{1 - \frac{\sum_{i=1}^N (\bar{y}_i - \bar{y})^2}{\sum_{i=1}^N (\bar{y}_i - \bar{y}_i)^2}} .$$

Якщо R близько до одиниці, то рівняння регресії добре описує експериментальні дані.

7. Виконати звіт на тему завдання згідно варіанта, умову, а також відповідні розрахунки та лістинг програмного коду виконання завдання. Повинні бути описані дії та хід виконання роботи. Після закінчення виконання завдання, по отриманим даним зробити висновки (детальний опис дивитися в завданні).

В табл.2.2.3 (варіант № 1-10) по варіантам наведено плани і результати експерименту. Завдання обираються по варіантам з журналу обліку успішності, або їх задає викладач. Таблиця 2.2.3.- Варіант завдань по варіантам 1-10.

Варіант №1

№ крапок плану	Фактори								Параметри оптимізації		
	b ₀ , j=0	x ₁ , j=1	x ₂ , j=2	x ₃ , j=3	x ₁ x ₂ , j=4	x ₂ x ₃ , j=5	x ₃ x ₁ , j=6	x ₁ x ₂ x ₃ , j=7	y' _i , p=1	y'' _i , p=2	$\bar{y}_i$
1	+	+	+	+	+	+	+	+	34	38	
2	+	-	+	+	-	+	-	-	-2	-4	
3	+	+	-	+	-	-	+	-	-18	-24	
4	+	-	-	+	+	-	-	+	32	40	
5	+	+	+	-	+	-	-	-	-12	-16	
6	+	-	+	-	-	-	+	+	-2	-6	
7	+	+	-	-	-	+	-	+	-26	-32	
8	+	-	-	-	+	+	+	-	-8	-10	

Варіант №2

№ крапок плану	Фактори								Параметри оптимізації		
	$b_0,$ j=0	$x_1,$ j=1	$x_2,$ j=2	$x_3,$ j=3	$x_1x_2,$ j=4	$x_2x_3,$ j=5	$x_3x_1,$ j=6	$x_1x_2x_3,$ j=7	$y_i',$ p=1	$y_i'',$ p=2	$\bar{y}_i$
1	+	+	+	+	+	+	+	+	95	89	
2	+	-	+	+	-	+	-	-	7	8	
3	+	+	-	+	-	-	+	-	9	6	
4	+	-	-	+	+	-	-	+	-1	2	
5	+	+	+	-	+	-	-	-	15	11	
6	+	-	+	-	-	-	+	+	-3	-4	
7	+	+	-	-	-	+	-	+	0	3	
8	+	-	-	-	+	+	+	-	0	-4	

Варіант №3

№ крапок плану	Фактори								Параметри оптимізації		
	$b_0,$ j=0	$x_1,$ j=1	$x_2,$ j=2	$x_3,$ j=3	$x_1x_2,$ j=4	$x_2x_3,$ j=5	$x_3x_1,$ j=6	$x_1x_2x_3,$ j=7	$y_i',$ p=1	$y_i'',$ p=2	$\bar{y}_i$
1	+	+	+	+	+	+	+	+	95	98	
2	+	-	+	+	-	+	-	-	18	21	
3	+	+	-	+	-	-	+	-	17	21	
4	+	-	-	+	+	-	-	+	-1	-2	
5	+	+	+	-	+	-	-	-	30	36	
6	+	-	+	-	-	-	+	+	-3	-5	
7	+	+	-	-	-	+	-	+	4	6	
8	+	-	-	-	+	+	+	-	-8	-10	

Варіант №4

№ крапок плану	Фактори								Параметри оптимізації		
	b ₀ , j=0	x ₁ , j=1	x ₂ , j=2	x ₃ , j=3	x ₁ x ₂ , j=4	x ₂ x ₃ , j=5	x ₃ x ₁ , j=6	x ₁ x ₂ x ₃ , j=7	y' _i , p=1	y'' _i , p=2	$\bar{y}_i$
1	+	+	+	+	+	+	+	+	148	150	
2	+	-	+	+	-	+	-	-	-100	-110	
3	+	+	-	+	-	-	+	-	-116	-104	
4	+	-	-	+	+	-	-	+	104	103	
5	+	+	+	-	+	-	-	-	-110	-114	
6	+	-	+	-	-	-	+	+	89	99	
7	+	+	-	-	-	+	-	+	88	98	
8	+	-	-	-	+	+	+	-	-100	-90	

Варіант №5

№ крапок плану	Фактори								Параметри оптимізації		
	b ₀ , j=0	x ₁ , j=1	x ₂ , j=2	x ₃ , j=3	x ₁ x ₂ , j=4	x ₂ x ₃ , j=5	x ₃ x ₁ , j=6	x ₁ x ₂ x ₃ , j=7	y' _i , p=1	y'' _i , p=2	$\bar{y}_i$
1	+	+	+	+	+	+	+	+	19	22	
2	+	-	+	+	-	+	-	-	2	3	
3	+	+	-	+	-	-	+	-	-2	-3	
4	+	-	-	+	+	-	-	+	1	0	
5	+	+	+	-	+	-	-	-	1	2	
6	+	-	+	-	-	-	+	+	-8	-5	
7	+	+	-	-	-	+	-	+	-7	-6	
8	+	-	-	-	+	+	+	-	1	1	

Варіант №6

№ крапок плану	Фактори								Параметри оптимізації		
	b_0 , j=0	x_1 , j=1	x_2 , j=2	x_3 , j=3	x_1x_2 , j=4	x_2x_3 , j=5	x_3x_1 , j=6	$x_1x_2x_3$, j=7	y'_i , p=1	y''_i , p=2	$\bar{y}_i$
1	+	+	+	+	+	+	+	+	42	39	
2	+	-	+	+	-	+	-	-	4	2	
3	+	+	-	+	-	-	+	-	-1	-3	
4	+	-	-	+	+	-	-	+	22	21	
5	+	+	+	-	+	-	-	-	1	-2	
6	+	-	+	-	-	-	+	+	-3	-4	
7	+	+	-	-	-	+	-	+	-4	-5	
8	+	-	-	-	+	+	+	-	1	1	

Варіант №7

№ крапок плану	Фактори								Параметри оптимізації		
	b_0 , j=0	x_1 , j=1	x_2 , j=2	x_3 , j=3	x_1x_2 , j=4	x_2x_3 , j=5	x_3x_1 , j=6	$x_1x_2x_3$, j=7	y'_i , p=1	y''_i , p=2	$\bar{y}_i$
1	+	+	+	+	+	+	+	+	35	39	
2	+	-	+	+	-	+	-	-	2	4	
3	+	+	-	+	-	-	+	-	-2	-3	
4	+	-	-	+	+	-	-	+	15	18	
5	+	+	+	-	+	-	-	-	1	0	
6	+	-	+	-	-	-	+	+	9	6	
7	+	+	-	-	-	+	-	+	7	5	
8	+	-	-	-	+	+	+	-	-1	-1	

Варіант №8

№ крапок плану	Фактори								Параметри оптимізації		
	$b_0,$ j=0	$x_1,$ j=1	$x_2,$ j=2	$x_3,$ j=3	$x_1x_2,$ j=4	$x_2x_3,$ j=5	$x_3x_1,$ j=6	$x_1x_2x_3,$ j=7	$y'_i,$ p=1	$y''_i,$ p=2	$\bar{y}_i$
1	+	+	+	+	+	+	+	+	2	-1	
2	+	-	+	+	-	+	-	-	-10	-9	
3	+	+	-	+	-	-	+	-	-10	-11	
4	+	-	-	+	+	-	-	+	2	0	
5	+	+	+	-	+	-	-	-	10	9	
6	+	-	+	-	-	-	+	+	12	14	
7	+	+	-	-	-	+	-	+	-10	-9	
8	+	-	-	-	+	+	+	-	6	4	

Варіант №9

№ крапок плану	Фактори								Параметри оптимізації		
	$b_0,$ j=0	$x_1,$ j=1	$x_2,$ j=2	$x_3,$ j=3	$x_1x_2,$ j=4	$x_2x_3,$ j=5	$x_3x_1,$ j=6	$x_1x_2x_3,$ j=7	$y'_i,$ p=1	$y''_i,$ p=2	$\bar{y}_i$
1	+	+	+	+	+	+	+	+	24	26	
2	+	-	+	+	-	+	-	-	8	10	
3	+	+	-	+	-	-	+	-	10	13	
4	+	-	-	+	+	-	-	+	10	15	
5	+	+	+	-	+	-	-	-	11	16	
6	+	-	+	-	-	-	+	+	4	7	
7	+	+	-	-	-	+	-	+	0	-3	
8	+	-	-	-	+	+	+	-	8	7	

Варіант №10

№ крапок плану	Фактори								Параметри оптимізації		
	$b_0,$ j=0	$x_1,$ j=1	$x_2,$ j=2	$x_3,$ j=3	$x_1x_2,$ j=4	$x_2x_3,$ j=5	$x_3x_1,$ j=6	$x_1x_2x_3,$ j=7	$y'_i,$ p=1	$y''_i,$ p=2	$\bar{y}_i$
1	+	+	+	+	+	+	+	+	100	110	
2	+	-	+	+	-	+	-	-	76	80	
3	+	+	-	+	-	-	+	-	25	29	
4	+	-	-	+	+	-	-	+	34	40	
5	+	+	+	-	+	-	-	-	-2	-4	
6	+	-	+	-	-	-	+	+	21	26	
7	+	+	-	-	-	+	-	+	-43	-50	
8	+	-	-	-	+	+	+	-	-65	-70	

Питання для самостійної роботи

1. Поясніть головну мету завдання оптимізації.
2. Поясніть поняття фактору експериментів.
3. Що значить повний факторний експеримент?
4. Поясніть властивості ортогональності, симетричності та нормованості.
5. Поясніть поняття адекватності.
6. Для чого використовується функція відгуку?

Рекомендована література

[1], [2], [4], [5].

2.3 ПРАКТИЧНА РОБОТА №3 - ДРОБОВИЙ ФАКТОРНИЙ ЕКСПЕРИМЕНТ

Основні теоретичні положення.

Метою роботи є набуття практичних навичок у застосуванні плану дробового факторного експерименту (ДФЕ) для побудови математичних моделей об'єктів управління та процесів, а також освоєння методики статистичного аналізу отриманих рівнянь регресії.

Ідея ДФЕ полягає в тому, щоб скоротити кількість дослідів ПФЕ та щоб матриця планування зберегла властивість ортогональності [4]. На початкових етапах досліджень потрібно отримати деяку, не дуже точну інформацію про процес при мінімальних витратах на проведення експерименту. Використання дрібних реплік від повного факторного експерименту дозволяє зменшити кількість дослідів. Якщо вплив деякої взаємодії факторів визнається малим, то коефіцієнт регресії за такої взаємодії виявляється незначним. Це дозволяє використовувати відповідний стовпець матриці планування ПФЕ з метою оцінки впливу додаткового фактора.

Припустимо, що слід досліджувати вплив на процес трьох чинників і отримати математичне опис у вигляді лінійного рівняння. Відомо, що взаємодія факторів x_1 та x_2 є незалежною. Цим можна скористатися при вивченні впливу на процес третього чинника x_3 . Виберемо матрицю повного факторного експерименту 2^2 і прирівняємо добуток x_1x_2 до фактору x_3 . При цьому планування 2^{3-1} буде представлено генеруючим співвідношенням: $x_3 = x_1x_2$ (табл.2.3.1).

Помножуючи генеруюче співвідношення на x_3 отримуємо визначальний контраст:

$$1=x_1x_2x_3,$$

$$x_3^2=1.$$

Таблиця 2.3.1 - Планування експерименту типу 2^{3-1}

Досліди	x_0	x_1	x_2	x_1x_2	x_3	Залежні зміни
1	+1	-1	-1	+1	+1	y_1
2	+1	-1	+1	-1	-1	y_2
3	+1	+1	-1	-1	-1	y_3
4	+1	+1	+1	+1	+1	y_4

Контраст допомагає визначати змішані ефекти. Щоб визначити, який ефект змішаний з даним, потрібно помножити обидві частини визначального контраста на стовпець, відповідний даному ефекту. Отримане співвідношення називається генеруючим. Генеруюче співвідношенням, показує, з яким із ефектів змішаний цей ефект.

Для x_1 маємо: $x_1 = x_1^2 x_2 x_3 = x_2 x_3$,

для x_2 знаходимо: $x_2 = x_1 x_2^2 x_3 = x_1 x_3$,

для x_3 : $x_3 = x_1 x_2$,

для x_0 : $x_0 = x_1 x_2 x_3$.

Це означає, що коефіцієнти лінійного рівняння будуть оцінками:

$$b_1 \rightarrow \beta_1 + \beta_{23},$$

$$b_2 \rightarrow \beta_2 + \beta_{13},$$

$$b_3 \rightarrow \beta_3 + \beta_{12},$$

$$b_0 \rightarrow \beta_0 + \beta_{123},$$

де β_i - Справжні коефіцієнти регресії;

b_i – оцінки коефіцієнтів регресії, обчислені за даними вибірки.

Зі зростанням числа факторних змінних збільшується дрібність реплік та ускладняється система змішування. ДФЕ не дозволяє окремо оцінювати ефекти факторів та взаємодій, та отримують змішані оцінки, що є платою за зменшення кількості дослідів у матриці планування.

Приклади розв'язання

Приклад 1.

Досліджується хімічний процес, що протікає в реакторі. Як функція відгуку «у» розглядається вихід готового продукту. Як впливають Розглядається вплив п'яти факторів x_1 , x_2 , x_3 , x_4 , x_5 . Для побудови математичної моделі процесу використовували дробовий факторний експеримент 2^{5-2} з генеруючими співвідношеннями $x_4 = x_1x_2x_3$, $x_5 = -x_1x_2$. При цьому передбачалося, що ефекти взаємодії $x_1x_2x_3$, x_1x_2 незначні. Вони замінені в матриці планування на x_4 і x_5 (табл.2.3.2).

Таблиця 2.3.2 – Матриця планування та результати ДФЕ 2^{5-2}

Досліди	x_0	x_1	x_2	x_3	$x_4 = x_1x_2x_3$	$x_5 = -x_2x_3$	у
1	+1	-1	-1	-1	-1	-1	14,5
2	+1	+1	+1	-1	-1	-1	18,6
3	+1	-1	-1	+1	+1	-1	13,0
4	+1	+1	-1	+1	-1	+1	51
5	+1	-1	+1	+1	-1	+1	23,2
6	+1.	+1	-1	-1	+1	+1	41
7	+1	-1	+1	-1	+1	+1	38
8	+1	+1	+1	+1	+1	-1	17,6

Визначимо систему спільних оцінок коефіцієнтів регресії. Визначальний контраст має вигляд:

$$1 = x_1x_2x_3x_4;$$

$$1 = -x_1 x_2 x_5.$$

Перемноживши перші два вирази отримуємо: $1 = -x_3 x_4 x_5$.

Узагальнюючий визначальний контраст має вигляд:

$$1 = x_1 x_2 x_3 x_4 = -x_1 x_2 x_5 = -x_3 x_4 x_5.$$

Помножуючи кожен із чинників на узагальнюючий контраст отримуємо систему оцінок коефіцієнтів регресії:

$$b_1 \rightarrow \beta_1 + \beta_{234} - \beta_{25} - \beta_{1345};$$

$$b_2 \rightarrow \beta_2 + \beta_{134} - \beta_{15} - \beta_{2345};$$

$$b_3 \rightarrow \beta_3 + \beta_{124} - \beta_{1235} - \beta_{45};$$

$$b_4 \rightarrow \beta_4 + \beta_{123} - \beta_{1245} - \beta_{35};$$

$$b_5 \rightarrow \beta_5 + \beta_{12345} - \beta_{12} - \beta_{34}.$$

Розрахуємо значення коефіцієнтів регресії з використанням рівняння:

$$b_j = \frac{1}{N} \sum_{i=1}^N x_{ji} y_i, \quad j = 0, 1, \dots, k$$

Отримаємо такі результати: $b_0 = 27,22$; $b_1 = 4,83$; $b_2 = -2,85$; $b_3 = -0,83$; $b_4 = -0,4$; $b_5 = 11,1$.

Перевіримо значущість коефіцієнтів регресії знаючи величину дисперсії відтворюваності дослідів, що дорівнює $s^2_{\text{восп}} = 2,81$ з числом ступенів свободи $f = 2$.

Оцінка дисперсії похибки коефіцієнтів регресії дорівнює:

$$s^2_{bi} = s^2_{\text{восп}} / N = 2,81 / 8 = 0,35;$$

$$s_{bi} = 0,59.$$

Значимість коефіцієнтів регресії перевіряємо з використанням критерію Стюдента:

$$t_i = \frac{|b_i|}{\sqrt{s_{bi}^2}}.$$

Табличне значення критерію рівня значимості $\alpha=0,05$ і числа ступенів свободи $f = 2$ дорівнює 4,3 (додаток А). Розрахункові значення критерію дорівнюють:

$$t_{b1} = 4,83 / 0,59 = 8,2 > 4,3 \text{ коефіцієнт значимий};$$

$$t_{b2} = 4,83 > 4,3 \text{ коефіцієнт значимий};$$

$$t_{b3} = 1,4 < 4,3 \text{ коефіцієнт не значущий};$$

$$t_{b4} = 0,7 < 4,3 \text{ коефіцієнт не значущий};$$

$$t_{b5} = 18,8 > 4,4 \text{ коефіцієнт значущий};$$

Перевіряємо адекватність отриманого рівняння регресії за критерієм Фішера:

$$F = \frac{S_{ост}^2}{S_{восп}^2},$$

Розраховані за моделлю $y_{\text{мод}} = 27,22 + 4,83x_1 - 2,85x_2 + 11,1x_5$

значення виходу готового продукту наведені в табл.2.3.3.

.

Таблиця 2.3.3 – Розрахункові значення виходу товару

Дослідження	1	2	3	4	5	6	7	8
$y_{\text{мод}}$	14,14	18,1	14,14	46	30,64	46	30,64	18,1

Залишкова дисперсія моделі дорівнює:

$$S_{ост}^2 = \frac{1}{(N-k-1)} \sum_{i=1}^N (y_i - y_{имод})^2 = 40,36.$$

Розрахункове значення критерію Фішера менше табличного значення 19,25, яке отримано для рівня значимості $\alpha=0,05$, числа ступенів свободи $f_1=4$ та $f_2=2$:

$$F = \frac{S_{ост}^2}{S_{восп}^2} = \frac{40,43}{2,81} = 14,55 < F_m = 19,25.$$

Рівняння регресії адекватно експериментальним даним та може використовуватися для реалізації руху у область оптимуму.

Перейдемо від кодованих значень факторних змінних до фізичних змінних з урахуванням співвідношень:

$$x_1 = (z_1 - 2) / 1 = z_1 - 2;$$

$$x_2 = (z_2 - 2) / 10 = 0,1z_2 - 10;$$

$$x_5 = (z_5 - 2) / 1 = z_5 - 2.$$

$$y = 23,86 + 4,83 z_1 - 0,285 z_2 + 11,1 z_5.$$

Використовуючи рівняння у фізичних змінних, дослідник здійснює пошук оптимальних умов ведення процесу.

Приклад 2.

Дано матрицю планування ДФЕ і результати експерименту (табл. 2.3.4). Дисперсія відтворюваності дослідів дорівнює $s_{восп}^2 = 1,35$ за числа ступенів свободи $f = 2$. Було обрано напіврепліку 2^{4-1} із генерувальним співвідношенням $x_4 = -x_1 x_2$. Матрицю планування 2^{4-1} з результатами обчислень наведено в табл. 2.3.4.

Таблиця 2.3.4 – Матриця планування 2^{4-1} з результатами обчислень

	z ₁	z ₂	z ₃	z ₄		
Основний рівень	7,5	725	65	4		
Інтервал варіювання	1,0	25	10	1		
Верхній рівень	8,5	750	75	5		
Нижний рівень	6,5	700	55	3		
Досліди:	x0	x1	x2	x3	x4	y
1	+1	-1	-1	-1	-1	100
2	+1	+1	-1	-1	+1	98
3	+1	-1	+1	-1	+1	90
4	+1	+1	+1	-1	-1	90
5	+1	-1	-1	+1	-1	95
6	+1	+1	-1	+1	+1	86
7	+1	-1	+1	+1	+1	100
8	+1	+1	+1	+1	-1	99,5

Треба застосувати план дробового факторного експерименту (ДФЕ) для побудови математичних моделей об'єктів управління та процесів, а також освоєння методики статистичного аналізу отриманих рівнянь регресії.

В результаті виконання треба:

- 1) Скласти матрицю планування ДФЕ з урахуванням обраних генерувальних співвідношень;
- 2) Записати систему спільних оцінок коефіцієнтів регресії;
- 3) Визначити дисперсію похибки обчислення коефіцієнтів регресії;
- 4) Оцінити значущість розрахункових коефіцієнтів регресії за допомогою t-критерію Ст'юдента. Відкинути незначущі коефіцієнти зі структури моделі;
- 5) Розрахувати залишкову дисперсію похибки моделі;
- 6) Перевірити адекватність рівняння регресії за F-критерієм Фішера;
- 7) Проаналізувати вплив факторів на залежну змінну.

Розв'язування

Для виконання роботи скористаємося мовою програмування Python. Для початку імпортуємо необхідні модулі.

```
import numpy as np
```

```
import pandas as pd
import statsmodels.api as sm
import math
```

Тепер складемо матрицю планування ДФЕ з урахуванням обраних генерувальних співвідношень

Визначаємо матрицю планування ДФЕ з урахуванням обраних генерувальних співвідношень

```
X = np.array([[1, 1, -1, -1, -1],
              [1, 1, -1, -1, 1],
              [1, -1, 1, -1, 1],
              [1, 1, 1, -1, -1],
              [1, -1, -1, 1, -1],
              [1, 1, -1, 1, 1],
              [1, -1, 1, 1, 1],
              [1, 1, 1, 1, -1]])
```

Визначаємо значення залежної змінної

```
y = np.array([100, 98, 90, 90, 95, 86, 100, 99.5])
```

Після цього потрібно записати систему спільних оцінок коефіцієнтів регресії. Результат розрахунку представлено на рис.2.3.1-рис.2.3.5.

Обчислити оцінки коефіцієнтів регресії за допомогою
методу OLS (Ordinary Least Squares) з бібліотеки statsmodels:

```
model = sm.OLS(y, X).fit()
```

```
beta_hat = model.params
```

Таким чином, система спільних оцінок коефіцієнтів

регресії матиме такий вигляд:

```
print(beta_hat)
```

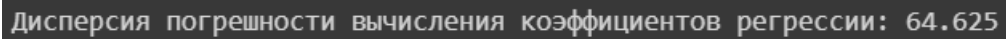
```
[ 9.49375e+01 -5.00000e-01 -6.25000e-02  1.87500e-01 -1.43750e+00]
```

Рисунок 2.3.1 – Результат розрахунку оцінок коефіцієнтів регресії

Тепер визначимо дисперсію похибки обчислення коефіцієнтів регресії.

```
mse_resid = model.mse_resid
```

```
print("Дисперсія похибки обчислення коефіцієнтів регресії:", mse_resid)
```



Дисперсия погрешности вычисления коэффициентов регрессии: 64.625

Рисунок 2.3.2 – Результат розрахунку дисперсії похибки

Далі оцінюємо значущість розрахункових коефіцієнтів регресії за допомогою t-критерію Стюдента. Відкинемо незначущі коефіцієнти зі структури моделі.

```
# Для оцінки значущості розрахункових коефіцієнтів регресії з
```

```
# допомогою t-критерію Стюдента в Python за допомогою бібліотеки
```

```
# statsmodels можна використовувати метод summary об'єкта model.
```

```
# виводимо таблицю коефіцієнтів регресії
```

```
print(model.summary())
```

```
# відкидаємо незначущі коефіцієнти зі структури моделі
```

```
p_values = model.pvalues
```

```
significant_vars_idx = np.where(p_values < 0.05)[0]
```

```
significant_results = model.params[significant_vars_idx]
```

```
# виводимо оцінки значущих коефіцієнтів регресії
```

```
print("Оцінки значущих коефіцієнтів регресії:")
```

```
print(significant_results)
```

OLS Regression Results						
=====						
Dep. Variable:	y	R-squared:	0.077			
Model:	OLS	Adj. R-squared:	-1.154			
Method:	Least Squares	F-statistic:	0.06226			
Date:	Sat, 08 Apr 2023	Prob (F-statistic):	0.989			
Time:	16:18:44	Log-Likelihood:	-24.103			
No. Observations:	8	AIC:	58.21			
Df Residuals:	3	BIC:	58.60			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	94.9375	2.958	32.092	0.000	85.523	104.352
x1	-0.5000	3.282	-0.152	0.889	-10.944	9.944
x2	-0.0625	2.958	-0.021	0.984	-9.477	9.352
x3	0.1875	2.958	0.063	0.953	-9.227	9.602
x4	-1.4375	2.958	-0.486	0.660	-10.852	7.977
=====						
Omnibus:	3.353	Durbin-Watson:	1.545			
Prob(Omnibus):	0.187	Jarque-Bera (JB):	0.948			
Skew:	-0.172	Prob(JB):	0.623			
Kurtosis:	1.349	Cond. No.	1.73			
=====						
Notes:						
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.						
Оценки значимых коэффициентов регрессии:						
[94.9375]						

Рисунок 2.3.3 – Результат відкидання незначущих коефіцієнтів регресії

Розрахуємо залишкову дисперсію похибки моделі.

розраховуємо залишкову дисперсію похибки моделі

residual_variance = model.scale

print("Залишкова дисперсія похибки моделі:", residual_variance)

Залишкова дисперсія похибки моделі: 64.625

Рисунок 2.3.4 – Результат розрахунку залишкової дисперсії похибки моделі

Далі перевіримо адекватність рівняння регресії за F-критерієм Фішера. Щоб перевірити адекватність рівняння регресії за F-критерієм Фішера, необхідно порівняти розраховане значення F-статистики з теоретичним

значенням F-розподілу для заданого рівня значущості та числа ступенів свободи.

```
from scipy.stats import f
```

```
# Обчислимо F-статистику, використовуючи атрибут fvalue моделі
```

```
f_statistic = model.fvalue
```

```
print("F-статистика моделі", f_statistic)
```

```
# Обчислимо ступені свободи для чисельника та знаменника F-
```

```
статистики, використовуючи атрибути моделі df_model та df_resid відповідно
```

```
dfn = model.df_model
```

```
dfd = model.df_resid
```

```
# Обчислимо критичне значення F-
```

```
розподілу за допомогою методу ppf з модуля scipy.stats
```

```
alpha = 0.05
```

```
f_critical = f.ppf(1 - alpha, dfn, dfd)
```

```
print("Критичне значення F-розподілу моделі", f_critical)
```

```
# Порівняємо розраховану F-статистику з критичним значенням F-розподілу.
```

```
# Якщо розрахована F-
```

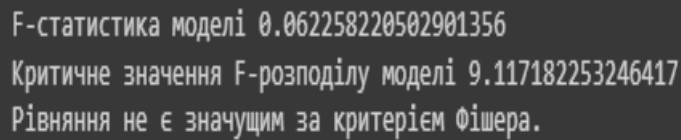
```
статистика більша за критичне значення, ми відкидаємо нульову гіпотезу і робимо висновок, що модель є значущою.
```

```
if f_statistic > f_critical:
```

```
    print("Рівняння регресії є значущим за критерієм Фішера.")
```

```
else:
```

```
    print("Рівняння не є значущим за критерієм Фішера.")
```



F-статистика моделі 0.062258220502901356
Критичне значення F-розподілу моделі 9.117182253246417
Рівняння не є значущим за критерієм Фішера.

Рисунок 2.3.5 – Результат перевірки адекватності рівняння регресії за критерієм Фішера

Для аналізу впливу факторів на залежну змінну можна поглянути на значення коефіцієнтів регресії. Знак коефіцієнта вказує на напрямок впливу фактора на залежну змінну: якщо коефіцієнт позитивний, то збільшення значення фактора призводить до збільшення значення залежної змінної, а якщо негативний - то до зменшення.

Виходячи зі значень коефіцієнтів регресії в нашій моделі, можна зробити такі висновки:

Фактор x_1 (лінійний член x_1) має позитивний вплив на y , тобто збільшення x_1 призводить до збільшення y .

Фактор x_2 (лінійний член x_2) має негативний вплив на y , тобто збільшення x_2 призводить до зменшення y .

Фактор x_3 (лінійний член x_3) чинить негативний вплив на y , тобто збільшення x_3 призводить до зменшення y .

Фактор x_4 (взаємодія між x_1 і x_2) має негативний вплив на y , тобто при збільшенні значення x_1 і x_2 одночасно, значення y зменшується. Також можна проаналізувати вплив факторів на y , побудувавши графіки залежності y від кожного фактора за фіксованих значень решти факторів. Це дасть змогу побачити, як змінюється y при зміні кожного фактора у відриві від інших.

Висновок: В результаті виконання даної практичної роботи набуто практичних навичок у застосуванні плану дробового факторного експерименту (ДФЕ) для побудови математичних моделей об'єктів управління та процесів, а також освоєння методики статистичного аналізу отриманих рівнянь регресії.

Завдання для виконання на практичному занятті

Варіант 1. Дана матриця планування ДФЕ та результати експерименту (табл. 2.3.5). Дисперсія відтворюваності дослідів дорівнює $s^2_{\text{восп}} = 1,35$ при числі ступенів свободи $f=2$. Була обрана напіврепліка 2^{4-1} з генеруючим співвідношенням $x_4 = -x_1x_2$. Матриця планування 2^{4-1} з результатами обчислень наведена у табл. 2.3.5.

Таблиця 2.3.5 - Матриця планування 2^{4-1} з результатами обчислень

	z_1	z_2	z_3	z_4	
Основний рівень	7,5	725	65	4	
Інтервал варіювання	1,0	25	10	1	
Верхній рівень	8,5	750	75	5	
нижній рівень	6,5	700	55	3	
Досліди:	x_0	x_1	x_2	x_3	x_4
1	+1	-1	-1	-1	-1
2	+1	+1	-1	-1	+1
3	+1	-1	+1	-1	+1
4	+1	+1	+1	-1	-1
5	+1	-1	-1	+1	-1
6	+1	+1	-1	+1	+1
7	+1	-1	+1	+1	+1
8	+1	+1	+1	+1	-1
					y
					100
					98
					90
					90
					95
					86
					100
					99,5

Розрахувати параметри лінійного рівняння регресії, провести статистичний аналіз отриманого рівняння.

Варіант 2 . Інтервали варіювання незалежних змінних z_1 - z_4 задані в таблиці 2.3.6. Відомо, що на залежну змінну «у» впливають дві парні взаємодії z_2z_3 і z_3z_4 . Вибрано генеруюче співвідношення $x_4 = x_1x_2$. Дисперсія відтворюваності дослідів дорівнює 400 зі ступенями свободи $f=2$.

Таблиця 2.3.6. - Інтервали варіювання факторних змінних

Параметри плану	z_1	z_2	z_3	z_4
z_0	0,87	40	1	250
Δz	0,15	5	0,25	50
$z_i=+1$	1,02	45	1,25	300
$z_i=-1$	0,72	35	0,75	200

Дано матрицю планування з результатами експериментів (табл. 2.3.7).

Таблиця 2.3.7. - Матриця ДФЕ 24-1 з результатами експерименту

Дослід	x_0	x_1	x_2	x_3	$x_4=x_1x_2$	y
1	+1	+1	+1	+1	+1	296
2	+1	+1	-1	+1	-1	122
3	+1	-1	-1	+1	+1	239
4	+1	-1	+1	+1	-1	586
5	+1	+1	+1	-1	+1	232
6	+1	+1	-1	-1	-1	292
7	+1	-1	-1	-1	+1	539
8	+1	-1	+1	-1	-1	383

Варіант 3. Матриця планування з результатами експерименту наведена у табл. 2.3.8. Досліджується процес очищення розсолу у виробництві хлору. Якість очищення характеризується залишковою концентрацією домішок іонів кальцію та магнію y_1 .

Як варіюються факторів обрані: співвідношення потоків α , дозування поліакриламідів $C_{\text{паа}}$, концентрація соди в освітленому розсолі $C_{\text{Na}_2\text{CO}_3}$, концентрація лугу в освітленому розсолі C_{NaOH} . З метою скорочення коштів та часу на проведення промислового експерименту планувався ДФЕ 2^{4-1} . Парна взаємодія факторів $-x_1x_2$ мало ймовірна, вона обрана як генеруюче співвідношення для фактора x_4 . Дисперсія відтворюваності оцінювалася за паралельними дослідями, становила величину $0,89 \text{ (мг/л)}^2$ при числі ступенів

свободи $f = 51$. Розрахувати лінійне рівняння регресії вихідної змінної y_1 , що характеризує залишкову концентрацію домішок іонів кальцію і магнію в освітленому розсолі.

Таблиця 2.3.8 - Матриця планування ДФЕ 2^{4-1} з результатами експерименту

Найменування		α	$C_{\text{паа}}$	$C_{\text{Na}_2\text{CO}_3}$	C_{NaOH}	Концентрація домішок
Основний рівень		0,57	2,15	0,63	0,73	мг/л
Верхній рівень +1		0,63	3,05	0,88	1,13	
Нижній рівень -1		0,51	1,25	0,38	0,33	
Досвіди	x_0	x_1	x_2	x_3	$x_4 = -x_1x_2$	y_1
1	+1	+1	-1	-1	+1	22,6
2	+1	-1	+1	+1	+1	6,45
3	+1	+1	-1	+1	+1	25,3
4	+1	-1	+1	-1	+1	7,2
5	+1	-1	-1	-1	-1	6,7
6	+1	+1	+1	+1	-1	3,79
7	+1	-1	-1	+1	-1	9,3
8	+1	+1	+1	-1	-1	7,2

Варіант 4. Досліджується процес очищення розсолу у виробництві хлору. Якість очищення характеризується залишковою концентрацією домішок іонів кальцію та магнію y_1 . Скласти рівняння регресії, що враховує парну взаємодію факторів x_2x_3 . Доповнити матрицю ДФЕ (табл. 2.3.8) стовпцем із парною взаємодією x_2x_3 .

Розрахувати рівняння регресії з урахуванням парної взаємодії x_2x_3 вихідної змінної y_1 , що характеризує залишкову концентрацію домішок іонів кальцію і магнію в освітленому розсолі.

Варіант 5. Досліджується процес очищення розсолу у виробництві хлору. Якість очищення характеризується залишковою концентрацією домішок іонів кальцію та магнію y_1 . Скласти рівняння регресії, що враховує парні взаємодії факторів. Доповнити матрицю ДФЕ (табл.2.3.8) стовпцем із парною взаємодією x_1x_3 , x_2x_3 .

Варіант 6. Дана матриця планування ДФЕ та результати експерименту (табл. 2.3.9). Дисперсія відтворюваності дослідів дорівнює $s^2_{\text{восп}}=1,35$ при числі ступенів свободи $f=2$. Була обрана напіврепліка 2^{4-1} з генеруючим співвідношенням $x_4=-x_1x_2$. Матриця планування 2^{4-1} з результатами обчислень наведена у табл. 2.3.9.

Таблиця 2.3.9 - Матриця планування 2^{4-1} з результатами обчислень

	z_1	z_2	z_3	z_4	
Основний рівень	7,5	725	65	4	
Інтервал варіювання	1,0	25	10	1	
Верхній рівень	8,5	750	75	5	
нижній рівень	6,5	700	55	3	
Досліди:	x_0	x_1	x_2	x_3	x_4
1	+1	-1	-1	-1	-1
2	+1	+1	-1	-1	+1
3	+1	-1	+1	-1	+1
4	+1	+1	+1	-1	-1
5	+1	-1	-1	+1	-1
6	+1	+1	-1	+1	+1
7	+1	-1	+1	+1	+1
8	+1	+1	+1	+1	-1
					y
					100
					98
					90
					90
					95
					86
					100
					99,5

Розрахувати параметри лінійного рівняння регресії, провести статистичний аналіз отриманого рівняння.

Варіант 6. Дано матрицю планування ДФЕ та результати експерименту (табл. 2.3.9). Дисперсія відтворюваності дослідів дорівнює $s^2_{\text{восп}}=1,35$ при числі ступенів свободи $f=2$. Була обрана напіврепліка 2^{4-1} з

генеруючим співвідношенням $x_4 = -x_1x_2$. Матриця планування 2^{4-1} з результатами обчислень наведена у табл. 2.3.9. Доповнить матрицю планування стовпцем з парною взаємодією x_1x_3 . Розрахувати рівняння регресії з урахуванням парної взаємодії x_1x_3 провести статистичний аналіз отриманого рівняння.

Варіант 7. Дано матрицю планування ДФЕ та результати експерименту (табл. 2.3.9). Дисперсія відтворюваності дослідів дорівнює $s^2_{\text{восп}} = 1,35$ при числі ступенів свободи $f=2$. Була обрана напіврепліка 2^{4-1} з генеруючим співвідношенням $x_4 = -x_1x_2$. Матриця планування 2^{4-1} з результатами обчислень наведена у табл. 2.3.9. Доповнить матрицю планування стовпцем з парною взаємодією x_2x_3 . Розрахувати рівняння регресії з урахуванням парної взаємодії x_2x_3 , провести статистичний аналіз отриманого рівняння.

Варіант 8. Дано матрицю планування ДФЕ та результати експерименту (табл. 2.3.9). Дисперсія відтворюваності дослідів дорівнює $s^2_{\text{восп}} = 1,35$ при числі ступенів свободи $f=2$. Була обрана напіврепліка 2^{4-1} з генеруючим співвідношенням $x_4 = -x_1x_2$. Матриця планування 2^{4-1} з результатами обчислень наведена у табл. 8. Доповнить матрицю планування стовпцями з парними взаємодією x_1x_3 та x_2x_3 . Розрахувати рівняння регресії з урахуванням парних взаємодій x_1x_3 та x_2x_3 , провести статистичний аналіз отриманого рівняння.

Варіант 9. Дана матриця планування ДФЕ та результати експерименту (табл. 2.3.9). Дисперсія відтворюваності дослідів дорівнює $s^2_{\text{восп}} = 1,35$ при числі ступенів свободи $f=2$. Була обрана напіврепліка 2^{4-1} з генеруючим співвідношенням $x_4 = -x_1x_2$. Матриця планування 2^{4-1} з результатами обчислень наведена у табл. 2.3.9. Доповнить матрицю планування стовпцями з парними взаємодією x_2x_3 та x_3x_4 . Розрахувати рівняння регресії з урахуванням парних взаємодій x_2x_3 та x_3x_4 , провести статистичний аналіз отриманого рівняння.

Варіант 10. Дана матриця планування ДФЕ та результати експерименту (табл. 2.3.9). Дисперсія відтворюваності дослідів дорівнює $s^2_{\text{восп}}=1,45$ при числі ступенів свободи $f=3$. Була обрана напіврепліка 2^{4-1} з генеруючим співвідношенням $x_4=-x_1x_2$. Матриця планування 2^{4-1} з результатами обчислень наведена у табл. 2.3.9. Доповнить матрицю планування стовпцями з парними взаємодією x_2x_3 та x_3x_4 . Розрахувати рівняння регресії з урахуванням парних взаємодій x_2x_3 та x_3x_4 , провести статистичний аналіз отриманого рівняння.

Питання для самостійної роботи

1. У чому суть дробового факторного експерименту?
2. Як будуватися матриця планування ДФЕ у кодованих змінних та фізичних змінних?
3. Що таке генеруюче співвідношення?
4. Як формується узагальнений визначальний контраст?
5. Як утворюється система спільних оцінок коефіцієнтів регресії?
6. У чому різниця планів ПФЕ та ДФЕ?

Рекомендована література

[1], [7], [8], [9].

2.4 ПРАКТИЧНА РОБОТА №4 - ОБРОБКА РЕЗУЛЬТАТІВ ПРАКТИЧНОГО ЕКСПЕРИМЕНТУ

Основні теоретичні положення.

Мета роботи: вивчення методики проведення експериментів за ортогональним планом, оволодіння теорією перевірки адекватності моделі оригіналу.

Знаходження відрядкової дисперсії

Припустимо, що в кожній точці факторного простору, якій відповідає один з рядків матриці планування, проводиться серія з m дослідів. Для будь-якої i -ої точки обчислюється середнє значення вихідної величини

$$y_i = \sum_{u=1}^m \frac{y_{iu}}{m} \quad (4.1.)$$

де m – число дослідів;

y_{iu} – значення вихідної величини у кожній строчці;

y_i – значення вихідної величини у кожній точці.

s^2 – відрядкову дисперсію вихідної величини:

$$s^2\{y_i\} = \sum_{u=1}^m \frac{(y_{iu} - y_i)^2}{m-1}. \quad (4.2)$$

Приклад 1.

Розглянемо етапи обробки результатів експерименту на основі результатів представлених в табл.2.4.1.

Таблиця 2.4.1 – Матриця планування

N	X_0	X_1	X_2	X_1X_2	Y_{1j}	Y_{2j}	Y_{3j}	Y_j	$S^2\{y_i\}$
1	+1	-1	-1	+1	43	35	48	42	43
2	+1	+1	-1	-1	90	86	94	90	16
3	+1	-1	+1	-1	10	16	16	14	12
4	+1	+1	+1	+1	56	54	58	56	4

Розв'язання

Середнє значення вихідної величини Y_i у кожній точці визначимо за формулою 4.1 ($m = 3$):

$$Y_1 = \frac{43 + 35 + 48}{3} = 42$$

$$Y_2 = \frac{90 + 86 + 94}{3} = 90$$

$$Y_3 = \frac{10 + 16 + 16}{3} = 14$$

$$Y_4 = \frac{56 + 54 + 58}{3} = 56$$

Визначимо підрядкову дисперсію за формулою 4.2:

$$S^2\{y_1\} = \frac{[(43 - 42)^2 + (35 - 42)^2 + (48 - 42)^2]}{2} = 43$$

$$S^2\{y_2\} = \frac{[(90 - 90)^2 + (86 - 90)^2 + (94 - 90)^2]}{2} = 43$$

$$S^2\{y_3\} = \frac{[(10 - 14)^2 + (16 - 14)^2 + (16 - 14)^2]}{2} = 43$$

$$S^2\{y_4\} = \frac{[(56 - 56)^2 + (54 - 56)^2 + (58 - 56)^2]}{2} = 43$$

Перевірка однорідності за критерієм Кохрена

Серед усієї сукупності розрахованих відрядкових дисперсій вибирається максимальна $S^2\{y_i\}_{\max}$ і береться відношення цієї дисперсії до

суми усіх відрядкових дисперсій $S^2\{y_i\}$, тобто визначається розрахункове значення коефіцієнта Кохрена, який показує, яку частку в загальній сумі відрядкових дисперсій має максимальна з них:

$$G_p = \frac{S^2\{y_i\} \max}{\sum_{i=1}^N S^2\{y_i\}},$$

У разі ідеальної однорідності відрядкових дисперсій коефіцієнт G_p прагнув би до значення $\frac{1}{N}$, де N – кількість дослідів (кількість рядків у матриці планування).

Розрахункове значення коефіцієнта Кохрена порівнюється з табличним значенням G критерію, яке вибирається з таблиць для прийнятого рівня значущості (i для чисел ступеня свободи відповідно чисельника f_1 і знаменника f_2 :

$$f_1 = m - 1; f_2 = N.$$

Для цього значення f_1 знаходиться в горизонтальному заголовку таблиці (вибирається стовпець), а f_2 вибирається ліворуч у вертикальному заголовку таблиці (вибирається рядок) і на перетині набуваємо табличного значення G_T коефіцієнта Кохрена. Якщо виконується умова $G_p < G_T$, то з вибраним рівнем статистичної значущості α (з достовірністю $1-\alpha$) усі підрядкові дисперсії визнаються однорідними. Інакше гіпотезу відкидають. За даними з нашого прикладу визначимо розрахункове значення коефіцієнта:

$$G_p = \frac{43}{43+16+12+4} = 0.57.$$

Відповідно до таблиці Б.1 (додаток Б) коефіцієнтів для $\alpha = 0,05$; $f_1 = 3 - 1 = 2$; $f_2 = 4$, знаходимо $G_p = 0.77$; $G_T > G_p$, тобто умова виконується.

Перевірка нуль-гіпотези за критерієм Стьюдента

Після перевірки однорідності переходять до визначення оцінок коефіцієнтів по формулі

$$a_k = \sum_{i=1}^N \frac{(y_{ik} X_{ik})}{N},$$

де k – номер вектор-стовпця.

У нашому прикладі маємо

$$a_0 = (42 + 90 + 14 + 56) / 4 = 50,5;$$

$$a_1 = (-42 + 90 - 14 + 56) / 4 = 22,5;$$

$$a_2 = (-42 - 90 + 14 + 56) / 4 = 15,5;$$

$$a_3 = (42 - 90 - 14 + 56) / 4 = 1,5.$$

Знайдені коефіцієнти рівняння регресії необхідно оцінити на статистичну значущість. Оцінка робиться за t-критерієм Стюдента. Для кожного коефіцієнта a_k обчислюється коефіцієнт (a_k – коефіцієнт рівняння регресії):

$$t_k = \frac{|a_k|}{S\{a_k\}},$$

тобто перевіряється відхилення від нуля знайденої оцінки. $S\{a_k\}$ – оцінка середнього квадратичного відхилення похибки визначення коефіцієнта.

Оцінка дисперсії коефіцієнтів, знайдених за експериментальними даними:

$$S^2\{a_k\} = \frac{S_{\epsilon}^2}{N \cdot m}$$
$$S\{a_k\} = \sqrt{S^2\{a_k\}}.$$

Оцінкою генеральної дисперсії відтворюваності S_B^2 , яка характеризує точність одного виміру, є середня з усіх відрядкових дисперсій:

$$S_B^2 = \sum_{i=1}^n \frac{S^2\{y_i\}}{N}.$$

За вибраного рівня статистичної значущості α (з таблиць розподілу Стьюдента за числом ступенів свободи $f = N(m - 1)$ знаходять табличне значення коефіцієнта $t_{\text{табл.}}$). Знайдене табличне значення порівнюється з розрахунковим значенням коефіцієнта. Якщо виконується нерівність $t_{\text{табл.}} > t_k$, то приймається нуль-гіпотеза, тобто вважається, що знайдений коефіцієнт a_k є статистично незначним і його слід виключити з рівняння регресії. Для цього прикладу:

$$S_B^2 = (43 + 16 + 12 + 4) / 4 = 18,75;$$

$$S\{a_k\} = \frac{18,75}{4 \cdot 3} = 1,25.$$

Визначимо розрахункові значення коефіцієнта Стьюдента:

$$t_0 = \frac{50,5}{1,25} = 40,4;$$

$$t_1 = \frac{22,5}{1,25} = 18;$$

$$t_2 = \frac{15,5}{1,25} = 12,4;$$

$$t_{12} = \frac{1,5}{1,25} = 1,2.$$

З таблиць за рівнем статистичної значущості $\alpha = 0,05$ і числом ступенів свободи $f = 4(3 - 1) = 8$ визначимо табличне значення коефіцієнта Стьюдента (додаток А).

Воно дорівнює $t_T = 2,3$.

Зіставимо розрахункові значення t_k з табличним t_T .

Нерівність виконується для t_{12} . Отже, можна припустити, що a_{12} статистично незначущий і його можна виключити з рівняння регресії.

Рівняння регресії, що містить статистично значимі коефіцієнти, буде (у кодованій системі):

$$Y' = 50.5 + 22.5x_1 - 15.5x_2.$$

Перевірка адекватності за критерієм Фішера

Отримане рівняння регресії необхідно перевірити на адекватність досліджуваному об'єкту. Для цієї мети необхідно оцінити, наскільки відрізняються середні значення y_i вихідної величини, отриманої в точках факторного простору, і значення y'_i , отриманого з рівняння регресії в тих же точках факторного простору. Для цього використовують дисперсію адекватності:

$$S_{ad}^2 = \frac{m}{N-L} \sum_{i=1}^n (y_i - y'_i)^2,$$

де L – число значущих коефіцієнтів.

Адекватність моделі перевіряють за F -критерієм Фішера:

$$F_p = \frac{S_{ad}^2}{S} * 2_v.$$

Знайдене розрахунковим шляхом F_p порівнюють з табличним значенням F_T (додаток В), яке визначається за рівнем значущості й числом ступенів свободи

$$f_{ад} = N - f_v = N(m - 1).$$

Якщо $F_p < F_T$, то отримана математична модель з прийнятим рівнем статистичної значущості α адекватна експериментальним даним.

Для цього прикладу отримуємо:

$$Y_1' = 50,5 + 22,5(-1) - 15,5(-1) = 43,5;$$

$$Y_2' = 50,5 + 22,5(+1) - 15,5(-1) = 88,5;$$

$$Y_3' = 50,5 + 22,5(-1) - 15,5(+1) = 12,5;$$

$$Y_4' = 50,5 + 22,5(+1) - 15,5(+1) = 57,5.$$

Розрахуємо оцінку дисперсії адекватності:

$$S_{ad}^2 = \frac{3[(42 - 43,5) \cdot 2 + (90 - 88,5) \cdot 2 + (14 - 12,5) \cdot 2 + (56 - 57,5) \cdot 2]}{4 - 3} = 27;$$

$$F_p = \frac{S_{ad}^2}{S_e^2} = \frac{27}{18,75} = 1,44.$$

Табличне значення коефіцієнта Фішера за рівнем статистичної значущості $\alpha = 0,05$ і числом ступенів свободи

$$f_{ад} = (4 - 3) = 1; f_{в} = 4(3 - 1) = 8$$

буде $F_T = 5,32$.

Отже, за вибраним рівнем статистичної значущості отримана в результаті експерименту

$$y' = 50,5 + 22,5x_1 - 15,5x_2$$

Модель адекватна досліджуваному об'єкту.

Приклади розв'язання

Приклад 2

Провести вивчення методики проведення експериментів за ортогональним планом, оволодіти теорією перевірки адекватності моделі оригіналу. В ході виконання роботи провести експеримент, повторивши 5 разів досліди в усіх точках факторного простору.

3,004	3,031	3,035	3,039	3,001
5,193	5,152	5,177	5,209	5,151
3,927	3,950	3,936	3,898	3,897
7,141	7,099	7,111	7,138	7,097

В роботі необхідно:

- 1) Скласти матрицю планування для повного трифакторного експерименту з використанням додаткового нульового фактора ($X_0 = 1$).
- 2) Перевірити однорідність дисперсії за критерієм Кохрена.
- 3) Знайти коефіцієнти рівняння регресії.
- 4) За допомогою критерію Ст'юдента оцінити значущість коефіцієнтів регресії.
- 5) Скласти рівняння регресії в кодованому вигляді й перевірити його адекватність за допомогою критерію Фішера.

Розв'язання

Для виконання роботи скористаємося мовою програмування Python. Використаємо модуль numpy для проведення розрахунків.

```
import numpy as np
```

```
x = np.array([[1, -1, -1, 1],  
              [1, 1, -1, -1],  
              [1, -1, 1, -1],  
              [1, 1, 1, 1]])
```

```
# Кількість дослідів (кількість рядків у матриці планування)
N=len(x)
print(N)
```

```
y=np.array([[3.004, 3.031, 3.035, 3.039, 3.001],
            [5.193, 5.125, 5.177, 5.209, 5.151],
            [3.927, 3.95, 3.936, 3.898, 3.897],
            [7.141, 7.099, 7.111, 7.138, 7.097]])
```

```
# Кількість дослідів
m=len(y[0])
print(m)
```

Кількість дослідів (кількість рядків у матриці планування) – 4.

Кількість дослідів – 5.

Після цього знайдемо середнє значення вихідної величини Y_i у кожній точці. Результати розрахунків представимо в табл. 2.4.2 - табл. 2.4.8.

```
y_mean=[]
for i in range(len(y)):
    y_mean.append(np.mean(y[i])/4)
print(y_mean)
```

```
[0.7555, 1.29275, 0.9804, 1.7792999999999999]
```

Рисунок 2.4.2 – Результат розрахунку середніх значень

Далі визначимо підрядкову дисперсію та перевіримо однорідність за критерієм Кохрена, серед усієї сукупності розрахованих відрядкових дисперсій вибирається максимальна і береться відношення цієї дисперсії до

суми усіх відрядкових дисперсій тобто визначається розрахункове значення коефіцієнта Кохрена, який показує, яку частку в загальній сумі відрядкових дисперсій має максимальна з них.

```
S_y=[]
for i in range(len(y_mean)):
    S_y.append(((y[i][0]-y_mean[i])**2+(y[i][1]-y_mean[i])**2+(y[i][2]-
y_mean[i])**2+(y[i][3]-y_mean[i])**2)/(m-1))
print(S_y)

Gp=max(S_y)/sum(S_y)
print(Gp)
```

```
[5.16103625, 15.080625562499998, 8.68723421, 28.547431389999996]
0.4966815500426612
```

Рисунок 2.4.3 – Результат розрахунку підрядкової дисперсії та перевірка однорідності за критерієм Кохрена

Тепер перевіримо нуль-гіпотези за критерієм Стюдента.

```
a=[]
for i in range(len(y)):
    a.append((x[i][0]*y_mean[0]+x[i][1]*y_mean[1]+x[i][2]*y_mean[2]+x[i][3]*y
_mean[3])/N)
print(a)
```

```
[0.06541249999999993, -0.17786249999999998, -0.3340375, 1.2019875]
```

Рисунок 2.4.4 – Результат визначення оцінок коефіцієнтів

Знайдені коефіцієнти рівняння регресії необхідно оцінити на статистичну значущість. Оцінка робиться за t-критерієм Стюдента. Для

кожного коефіцієнта a_k обчислюється коефіцієнт, тобто перевіряється відхилення від нуля знайденої оцінки. Для цього спочатку потрібно розрахувати оцінки дисперсії коефіцієнтів та оцінку генеральної дисперсії відтворюваності, яка характеризує точність одного виміру.

```
# Оцінка генеральної дисперсії
```

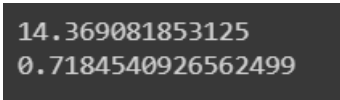
```
Sb=sum(S_y)/N
```

```
print(Sb)
```

```
# Оцінка дисперсії коефіцієнтів, знайдених за експериментальними даними
```

```
S_ak=Sb/(N*m)
```

```
print(S_ak)
```



```
14.369081853125  
0.7184540926562499
```

Рисунок 2.4.5 – Результат визначення генеральної дисперсії та дисперсії коефіцієнтів

Тепер ми можемо визначити значення коефіцієнта Стюдента.

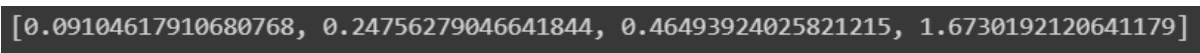
```
# Визначимо розрахункові значення коефіцієнта Стюдента
```

```
t=[]
```

```
for i in range(N):
```

```
    t.append(abs(a[i])/S_ak)
```

```
print(t)
```



```
[0.09104617910680768, 0.24756279046641844, 0.46493924025821215, 1.6730192120641179]
```

Рисунок 2.4.6 – Результат визначення коефіцієнта Стюдента

Рівняння регресії, що містить статистично значимі коефіцієнти, буде (у кодованій системі):

$$Y' = 0.065 - 0.1778x_1 - 0.334x_2 + 1.201x_3$$

Перевіримо адекватності за критерієм Фішера, отримане рівняння регресії необхідно перевірити на адекватність досліджуваному об'єкту. Для цієї мети необхідно оцінити, наскільки відрізняються середні значення вихідної величини, отриманої в точках факторного простору, і значення, отриманого з рівняння регресії в тих же точках факторного простору. Для цього використовують дисперсію адекватності.

```
y_i=[]
for i in range(len(y)):
    y_i.append(a[0]+a[1]*x[i][1]+a[2]*x[i][2]+a[3]*x[i][3])
print(y_i)
# Розраховуємо оцінку дисперсії адекватності
S_ad=(m*((y_mean[0]-y_i[0])*2+(y_mean[1]-y_i[1])*2+(y_mean[2]-
y_i[2])*2+(y_mean[3]-y_i[3])*2))/(1)
print(S_ad)
```

45.46300000000001

Рисунок 2.4.7 – Результат оцінки дисперсії адекватності

Тепер ми перевірити адекватність моделі за F-критерієм. Табличне значення коефіцієнта Фішера за рівнем статистичної значущості $\alpha = 0,05$ і числом ступенів свободи $f_{ad} = (4 - 3) = 1$; $f_b = 4(4 - 1) = 12$ буде $F_T = 4.75$.

```
# Перевіряємо адекватність моделі за F-критерієм
Fp=S_ad/Sb
print(Fp)
```

3.1639460659146206

Рисунок 2.4.8 – Результат перевірки адекватності моделі за F-критерієм.

Отже, за вибраним рівнем статистичної значущості отримана в результаті експерименту адекватна досліджуваному об'єкту.

Висновок: В результаті вивчення методики проведення експериментів за ортогональним планом, проведено ознайомлення з теорією перевірки адекватності моделі оригіналу. Розраховано генеральну дисперсію та дисперсії коефіцієнтів, знайдених за експериментальними даними. Перевірена нуль гіпотеза за критерієм Стюдента та перевірена адекватність моделі за F-критерієм.

Завдання для виконання на практичному занятті

Провести експеримент, повторивши 5 разів досліди в усіх точках факторного простору (знайти значення функції відгуку Y з табл.2.4.2 згідно з варіантом по журналу обліку .

Таблиця .2.4.2 – Вихідні значення функції відгуку во варіантах

Варіант 1					Варіант 2				
2,164	2,165	2,145	2,150	2,163	3,651	3,605	3,653	3,592	3,627
3,347	3,338	3,322	3,318	3,358	6,547	6,514	6,535	6,562	6,581
3,950	3,932	3,908	3,935	3,901	4,761	4,793	4,816	4,792	4,801
6,855	6,870	6,875	6,872	6,907	9,515	9,566	9,534	9,552	9,528
Варіант 3					Варіант 4				
2,124	2,150	2,139	2,140	2,157	2,588	2,597	2,542	2,537	2,539
3,382	3,394	3,368	3,374	3,372	4,191	4,165	4,152	4,129	4,138
2,705	2,652	2,655	2,674	2,713	3,201	3,231	3,202	3,199	3,248
4,307	4,242	4,276	4,317	4,255	5,509	5,453	5,448	5,511	5,445

Варіант 5	Варіант 6
3,072 3,028 3,080 3,049 3,069	4,292 4,285 4,333 4,304 4,277
5,193 5,159 5,163 5,220 5,168	8,385 8,390 8,404 8,421 8,390
3,932 3,955 3,893 3,915 3,939	5,881 5,886 5,847 5,900 5,909
7,094 7,126 7,149 7,102 7,158	13,349 13,332 13,357 13,342 13,356
Варіант 7	Варіант 8
4,307 4,284 4,284 4,316 4,286	3,583 3,605 3,623 3,623 3,587
8,387 8,396 8,430 8,389 8,404	6,555 6,564 6,523 6,559 6,511
5,832 5,873 5,856 5,843 5,862	4,795 4,790 4,776 4,798 4,744
13,329 13,304 13,328 13,340 13,312	9,504 9,530 9,524 9,557 9,530
Варіант 9	Варіант 10
3,054 3,032 3,024 3,046 3,019	2,549 2,537 2,563 2,564 2,569
5,147 5,170 5,178 5,190 5,177	4,118 4,164 4,155 4,126 4,151
3,926 3,895 3,937 3,931 3,915	3,236 3,220 3,202 3,212 3,207
7,117 7,121 7,101 7,130 7,091	5,445 5,485 5,449 5,472 5,455

1. Скласти матрицю планування для повного трифакторного експерименту з використанням додаткового нульового фактора ($X_0 = 1$).
2. Перевірити однорідність дисперсії за критерієм Кохрена.
3. Знайти коефіцієнти рівняння регресії.
4. За допомогою критерію Стюдента оцінити значущість коефіцієнтів регресії.
5. Скласти рівняння регресії в кодованому вигляді й перевірити його адекватність за допомогою критерію Фішера.
6. Навести лістинг програми та висновки по завданню

Питання для самостійної роботи

1. Опишіть план знаходження відрядкової дисперсії вихідної величини.

2. Для чого потрібне розрахункове значення коефіцієнта Кохрена і як він знаходиться?
3. Що таке критерій Ст'юдента і де він використовується?
4. Для чого оцінюють, наскільки відрізняються середні значення y_i вихідної величини, отриманої в точках факторного простору, і значення y_i , отриманого з рівняння регресії в тих же точках факторного простору?
5. Чим визначається F-критерій Фішера і як його застосовують?

Рекомендована література

[1],[3], [6], [10].

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Конспект лекцій з курсу «Планування і обробка результатів експерименту» (для студентів денної і заочної форм навчання спеціальності 141 – Електроенергетика, електротехніка та електромеханіка) / Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова ; уклад. : Л. А. Назаренко. – Харків : ХНУМГ ім. О. М. Бекетова, 2018. – 163 с.
2. Статюха, Г.О. Вступ до планування оптимального експерименту : навч. посіб. / Г.О. Статюха, Д.М. Складанний, О.С. Бонаренко. – К. : ІВЦ «Політехніка», 2011. – 117 с. – Режим доступу: https://ela.kpi.ua/bitstream/123456789/18103/1/Vstup_planyvanya_SSB.pdf .- (дата звернення: 22.03.2023.)
3. Планування і обробка даних наукового експерименту : консп. лекцій / В.В. Полтавець. – Донецьк : ДВНЗ ДонНТУ, 2008. – 52 с.
4. Засименко, В.М. Основи теорії планування експерименту : навч. посібник. — Львів : Вид-во ДУ «ЛП», – 2000. – 205 с.
5. Мотигін, В.В. Планування експерименту в інженерних дослідженнях (лабораторний практикум). Навчальний посібник. / ВВ. Мотигін, С.М. Павлов. – Вінниця : ВДТУ, 2001. – 82 с.
6. Ляшенко, Б. М. Методи обчислень : навч.–метод. посіб. для студ. фізико–математ. ф-ту / Б. М. Ляшенко, О. М. Кривонос, Т. А. Вакалюк. – Житомир : Вид–во ЖДУ, 2014. – 228 с.
7. Горват, А.А. Методи обробки експериментальних даних. – Ужгород : Вид-во УЖНУ «Говерла». – 2019 р. – 182 с. – http://www.dgma.donetsk.ua/docs/kafedry/avp/metod/ДПЕ_Навчальний_посібник_Горват.Pdf. – (дата звернення: 22.03.2023.)

8. Теорія планування експериментів: Виконання розрахунково-графічної роботи [Електронний ресурс] : навч. посіб. для студ. спеціальності 131 «Прикладна механіка», спеціалізації «Технологія машинобудування» / С.М. Лапач ; КПІ ім. Ігоря Сікорського. – Електронні текстові данні (1 файл: 3,31 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2020. – 86 с.
9. Шейко В.М. Організація та методика науково-дослідницької діяльності : підручник / В.М. Шейко , Н.М. Кушнарєнко . — 6-те вид., перероб. і допов. — К. : Знання, 2008. — 310 с. - Режим доступу: <https://chmnu.edu.ua/wp-content/uploads/2016/07/SHejko-V.M.-ta-inshi-Organizatsiya-tametodika-naukovo-doslidnitskoyi-diyalnosti.pdf>. - (дата звернення: 22.03.2023.)
10. Грищук, Ю.С. Основи наукових досліджень : навч. посіб. / Ю.С. Грищук. — Харків : НТУ «ХПІ», 2008. — 232 с. Режим доступу: <http://web.kpi.kharkov.ua/ea/wp-content/uploads/sites/25/2017/02/OND-Ukr.pdf>

ДОДАТОК А. КРИТЕРІЙ СТЬЮДЕНТА.

Таблиця А.1 – Рівні значущості $q = 0,001 \div 0,1$

Число ступенів вільності f	Рівень значущості q					
	0,1	0,05	0,02	0,01	0,002	0,001
1	6,314	12,71	31,82	63,66	318,3	636,6
2	2,920	4,303	6,965	9,925	22,33	31,60
3	2,353	3,182	4,541	5,841	10,21	12,92
4	2,132	2,776	3,747	4,604	7,173	8,610
5	2,015	2,571	3,365	4,032	5,894	6,869
6	1,943	2,447	3,143	3,707	5,208	5,959
7	1,895	2,365	2,998	3,499	4,785	5,408
8	1,860	2,306	2,896	3,355	4,501	5,041
9	1,833	2,262	2,821	3,250	4,297	4,781
10	1,812	2,228	2,764	3,169	4,144	4,587
11	1,796	2,201	2,718	3,106	4,025	4,437
12	1,782	2,179	2,681	3,055	3,930	4,318
13	1,771	2,160	2,650	3,012	3,852	4,221
14	1,761	2,145	2,624	2,977	3,787	4,140
15	1,753	2,131	2,602	2,947	3,733	4,073
16	1,746	2,120	2,583	2,921	3,686	4,015
17	1,740	2,110	2,567	2,898	3,646	3,965
18	1,734	2,101	2,552	2,878	3,610	3,922
19	1,729	2,093	2,539	2,861	3,579	3,883
20	1,725	2,086	2,528	2,845	3,552	3,850
21	1,721	2,080	2,518	2,831	3,527	3,819
22	1,717	2,074	2,508	2,819	3,505	3,792
23	1,714	2,069	2,500	2,807	3,485	3,768
24	1,711	2,064	2,492	2,797	3,467	3,745
25	1,708	2,060	2,485	2,787	3,450	3,725
26	1,706	2,056	2,479	2,779	3,435	3,707
27	1,703	2,052	2,473	2,771	3,421	3,689
28	1,701	2,048	2,467	2,763	3,408	3,674
29	1,699	2,045	2,462	2,756	3,396	3,660
30	1,697	2,042	2,457	2,750	3,385	3,646
40	1,684	2,021	2,423	2,704	3,307	3,551
50	1,676	2,009	2,403	2,678	3,261	3,496
75	1,665	1,992	2,377	2,643	3,202	3,425
100	1,660	1,984	2,364	2,626	3,174	3,390
∞	1,644	1,960	2,326	2,576	3,090	3,290

Таблиця А.2 – Коефіцієнт довіри t_{α} (квантилі розподілу Стюдента)

k	При α			k	При γ		
	0,95	0,99	0,999		0,95	0,99	0,999
1	12,706	63,657	636,619	19	2,093	2,861	3,883
2	4,303	9,925	31,598	20	2,086	2,845	3,850
3	3,182	5,841	12,941	21	2,080	2,831	3,819
4	2,776	4,604	8,610	22	2,074	2,819	3,792
5	2,571	4,032	6,859	23	2,069	2,807	3,797
6	2,447	3,707	5,959	24	2,064	2,797	3,745
7	2,365	3,499	5,405	25	2,060	2,787	3,725
8	2,306	3,355	5,041	26	2,056	2,779	3,707
9	2,262	3,250	4,781	27	2,052	2,763	3,674
10	2,23	3,169	4,587	28	2,048	2,763	3,674
11	2,201	3,106	4,437	29	2,045	2,756	3,659
12	2,179	3,055	4,318	30	2,042	2,750	3,646
13	2,160	3,012	4,221	40	2,021	2,704	3,551
14	2,145	2,977	4,140	50	2,008	2,677	3,497
15	2,131	2,947	4,073	60	2,000	2,660	3,460
16	2,120	2,921	4,015	80	1,990	2,639	3,416
17	2,110	2,898	3,965	100	1,984	2,626	3,391
18	2,101	2,878	3,922	∞	1,960	2,576	3,291

ДОДАТОК Б. КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ КОХРЕНА

Таблиця Б.1 Квантилі розподілу Кохрена G_{1-q} для $q=0,05$

k_1	k_2										
	1	2	3	4	5	6	8	10	16	36	∞
2	9985	9750	9392	9057	8772	8534	8159	7880	7341	6602	5000
3	9669	8709	7977	7454	7071	6771	6333	6025	5466	4748	3333
4	9065	7679	6841	6287	5895	5598	5175	4884	4366	3720	2500
5	8412	6838	5981	5441	5065	4783	4387	4118	3645	3066	2000
6	7808	6161	5321	4803	4447	4184	3817	3568	3135	2612	1667
7	7271	5612	4800	4307	3974	3726	3384	3154	2756	2278	1429
8	6798	5157	4377	3910	3595	3362	3043	2829	2462	2022	1250
9	6385	4775	4027	3584	3286	3067	2768	2568	2226	1820	1111
10	6020	4450	3733	3311	3029	2823	2541	2353	2032	1655	1000
12	5410	3924	3264	2880	2624	2439	2187	2020	1737	1403	0833
15	4709	3346	2758	2419	2195	2034	1815	1671	1429	1144	0667
20	3894	2705	2205	1921	1735	1602	1422	1303	1108	0879	0500
24	3434	2354	1907	1656	1493	1374	1216	1113	0942	0743	0417
30	2929	1980	1593	1377	1237	1137	1001	0921	0771	0604	0333
40	2370	1576	1259	1082	0968	0887	0780	0713	0595	0462	0250
60	1737	1131	0895	0765	0682	0623	0552	0497	0411	0316	0167
120	0998	0632	0495	0419	0371	0337	0292	0266	0218	0165	0083
∞	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000

Всі квантилі G_{1-q} менше одиниці, тому в таблиці приведені лише десяткові знаки після коми, перед якою у разі користування таблицею, потрібно ставити нуль цілих. Наприклад, при $k_1 = 6$, $k_2 = 3$ маємо $G_{0,95} = 0,5321$.

ДОДАТОК В. КРИТЕРІЙ ФІШЕРА

Таблиця В.1 – Квантилі розподілу Фішера F_{1-q} для $q=0,05$

k_2	k_1								
	1	2	3	4	5	6	12	24	∞
1	164,4	199,5	215,7	224,6	230,2	234,0	244,9	249	254,3
2	18,5	19,2	19,2	19,3	19,3	19,3	19,4	19,5	19,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,7	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	5,9	5,8	5,6
5	6,6	5,8	5,4	5,2	5,1	5,0	4,7	4,5	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,0	3,8	3,7
7	5,6	4,7	4,4	4,1	4,0	3,9	3,6	3,4	3,2
8	5,3	4,5	4,1	3,8	3,7	3,6	3,3	3,1	2,9
9	5,1	4,3	3,9	3,6	3,5	3,4	3,1	2,9	2,7
10	5,0	4,1	3,7	3,5	3,3	3,2	2,9	2,7	2,5
11	4,8	4,0	3,6	3,4	3,2	3,1	2,8	2,6	2,4
12	4,8	3,9	3,5	3,3	3,1	3,0	2,7	2,5	2,3
13	4,7	3,8	3,4	3,2	3,0	2,9	2,6	2,4	2,2
14	4,6	3,7	3,3	3,1	3,0	2,9	2,5	2,3	2,1
15	4,5	3,7	3,3	3,1	2,9	2,8	2,5	2,3	2,1
16	4,5	3,6	3,2	3,0	2,9	2,7	2,4	2,2	2,0
17	4,5	3,6	3,2	3,0	2,8	2,7	2,4	2,2	2,0
18	4,4	3,6	3,2	2,9	2,8	2,7	2,3	2,1	1,9
19	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
20	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
22	4,3	3,4	3,1	2,8	2,7	2,6	2,2	2,0	1,8
24	4,3	3,4	3,0	2,8	2,6	2,5	2,2	2,0	1,7
26	4,2	3,4	3,0	2,7	2,6	2,4	2,1	1,9	1,7
28	4,2	3,3	2,9	2,7	2,6	2,4	2,1	1,9	1,6
30	4,2	3,3	2,9	2,7	2,5	2,4	2,1	1,9	1,6
40	4,1	3,2	2,9	2,6	2,5	2,3	2,0	1,8	1,5
60	4,0	3,2	2,8	2,5	2,4	2,3	1,9	1,7	1,4
120	3,9	3,1	2,7	2,5	2,3	2,2	1,8	1,6	1,3
∞	3,8	3,0	2,6	2,4	2,2	2,1	1,8	1,5	1,0

ДОДАТОК Г. КРИТИЧНІ ТОЧКИ РОЗПОДІЛУ χ^2

Таблиця Г.1 Рівні значущості $q = 0,01 \div 0,1$

Число ступенів вільності f	Рівень значущості q			
	0,01	0,02	0,05	0,1
1	6,635	5,412	3,841	2,706
2	9,210	7,824	5,991	4,605
3	11,34	9,837	7,815	6,251
4	13,28	11,67	9,488	7,779
5	15,09	13,39	11,07	9,236
6	16,81	15,03	12,59	10,64
7	18,48	16,62	14,07	12,02
8	20,09	18,17	15,51	13,36
9	21,67	19,68	16,92	14,68
10	23,21	21,16	18,31	15,99
11	24,73	22,62	19,68	17,28
12	26,22	24,05	21,03	18,55
13	27,69	25,47	22,36	19,81
14	29,14	26,87	23,68	21,06
15	30,58	28,26	25,00	22,31
16	32,00	29,63	26,30	23,54
17	33,41	31,00	27,59	24,77
18	34,81	32,35	28,87	25,99
19	36,19	33,69	30,14	27,20
20	37,57	35,02	31,41	28,41
21	38,93	36,34	32,67	29,62
22	40,29	37,66	33,92	30,81
23	41,64	38,97	35,17	32,01
24	42,98	40,27	36,42	33,20
25	44,31	41,57	37,65	34,38

