Міністерство освіти і науки України
Національний університет водного господарства і природокористування
Навчально-науковий інститут
енергетики, автоматики та водного господарства
Кафедра гідротехнічного будівництва та гідравліки

01-04-98М

МЕТОДИЧНІ ВКАЗІВКИ

до практичних занять та самостійної роботи
з навчальної дисципліни

«Математичні методи і моделі ГТС»

для здобувачів вищої освіти першого (бакалаврського) рівня
за освітньо-професійною програмою
«Гідротехнічне будівництво, водна інженерія і водні технології»
спеціальності 194
«Гідротехнічне будівництво, водна інженерія і водні технології»
всіх форм навчання

Рекомендовано
науково-методичною радою з якості
ННІЕАВГ
Протокол № 6 від 28.01.2025 р.

Рівне – 2025

Методичні вказівки до практичних занять та самостійної роботи із навчальної дисципліни «Математичні методи і моделі ГТС» для здобувачів вищої освіти першого (бакалаврського) рівня за освітньо-професійною програмою «Гідротехнічне будівництво, водна інженерія і водні технології» спеціальності 194 «Гідротехнічне будівництво, водна інженерія і водні технології» всіх форм навчання. Частина 1. Статистична обробка даних та побудова математичних моделей в гідротехнічній галузі. [Електронне видання] / Дем'янюк А. В. – Рівне : НУВГП, 2024. – 37 с.

Укладач: Дем'янюк А. В., старший викладач кафедри гідротехнічного будівництва та гідравліки

Відповідальний за випуск – Волк Л. Р., в.о. завідувача кафедри гідротехнічного будівництва та гідравліки.

Керівник (гарант) Освітньо-професійної програми «Гідротехнічне будівництво» спеціальності 194 «Гідротехнічне будівництво, водна інженерія і водні технології» – Клімов С. В.

Попередня версія методичних вказівок – 041-52

Зміст

Вступ

Навчальна дисципліна «Математичні методи і моделі ГТС» розроблена для здобувачів вищої освіти першого (бакалаврського) рівня, що навчаються за освітньо-професійною програмою «Гідротехнічне будівництво, водна інженерія та водні технології» спеціальності 194 «Гідротехнічне будівництво, водна інженерія та водні технології».

Предметом навчальної дисципліни є методи та алгоритми статистичної обробки даних, моделювання на основі даних та чисельного розв’язання інженерних задач.

Метою викладання дисципліни «Математичні методи і моделі ГТС» є ознайомлення із основними принципами математичного моделювання, методами побудови математичних моделей, методами та критеріями статистичної обробки даних, принципами чисельного розв’язання математичних задач, що застосовуються при проектуванні гідротехнічних об’єктів.

Метою проведення практичних занять є закріплення та засвоєння теоретичного матеріалу з лекційного курсу. В результаті проведення практичних занять майбутні інженери повинні набути навички практичного застосовування методик статистичного аналізу даних; побудови статистичних та регресійних моделей; методів чисельного розв’язання задач в практиці розрахунків гідротехнічних об’єктів різного типу та призначення та використовувати сучасні програмні продукти для вирішення даних задач.

За навчальними планами освітньо-професійної програми «Гідротехнічне будівництво, водна інженерія та водні технології» навчальна дисципліна представлена лекційним курсом, практичними заняттями та самостійною роботою. Кількість кредитів ECTS – 5,0; загальний обсяг дисципліни складає 150 год.

Дані методичні вказівки містять опис структури навчальної дисципліни та методичні рекомендації щодо практичної частини Змістового модулю 1. Математичне моделювання при проектуванні гідротехнічних об’єктів.

В даних методичних вказівках наведено алгоритми використання засобів *MS Excel* для розв’язування задач.

1. Структура навчальної дисципліни

1.1. Перелік змістових модулів та тем лекцій і короткий зміст лекційного матеріалу

Змістовий модуль 1. Математичне моделювання при проектуванні гідротехнічних об'єктів

Тема 1. Застосування математичних методів і моделей в гідротехнічному будівництві

Поняття моделі і моделювання. Мета та задачі моделювання. Емпіричні і теоретичні задачі, спостереження і експеримент. Математичне моделювання як метод дослідження реальних об'єктів.

Тема 2. Основні принципи математичного моделювання

Властивості математичних моделей. Критерії істинності математичних моделей. Основні етапи побудови математичних моделей. Адекватність моделі. Статистичні гіпотези та статистичні критерії. Прогнозування як мета математичного моделювання. Використання можливостей *MS Excel* для побудови емпіричних формул. Лінії тренду, їх рівняння та ступінь достовірності. Статистичний аналіз даних за допомогою стандартних функцій *MS Excel*. Можливості та принципи застосування стандартного засобу Аналіз даних. Створення та опрацювання баз даних засобами *MS Excel*. Пошук даних, сортування і фільтри.

Тема 3. Моделювання функцій розподілу випадкових величин на основі вибіркових даних

Вибіркові дані. Перевірка даних на однорідність. Статистичні параметри вибірки. Закон розподілу випадкової величини. Аналітичний та емпіричний розподіл. Параметри закону розподілу. Моделюванні функцій розподілу випадкової величини на основі вибіркових даних. Використання нормального, логарифмічно-нормального законів розподілу, закону Вейбула-Гнеденко, розподілу Гумбеля і-го типу. Перевірка гіпотез щодо функції розподілу випадкової величини.

Тема 4. Види взаємозв'язків між величинами і кореляційний аналіз

Вивчення статистичного взаємозв'язку між величинами. Кореляція. Задачі кореляційного аналізу. Парна та множинна кореляція. Коефіцієнт кореляції Пірсона. Методи кореляційного аналізу. Інтерпретація результатів кореляційного аналізу. Гіпотеза про значущість оцінки коефіцієнта кореляції.

Тема 5. Основи аналізу часових рядів

Дані спостережень та часові ряди. Властивості рядів спостережень. Дані спостережень в гідротехнічному будівництві. Задачі аналізу рядів спостережень. Методи аналізу часових рядів. Сезонність в часових рядах. Методи прогнозування на основі часових рядів. Використання засобів середовища програмування *R program* для аналізу рядів даних.

Тема 6. Побудова регресійних моделей

Поняття регресії. Основи регресійного аналізу. Лінійна регресія. Багатофакторна лінійна регресія. Нелінійна регресія. Побудова регресійних моделей. Метод найменших квадратів. Метод середніх. Коефіцієнт детермінації. Критерії адекватності регресійної моделі.

Змістовий модуль 2. Математичні методи розв'язання інженерних задач

Тема 7. Чисельні методи розв'язання інженерних задач

Аналітичні та чисельні методи, їх переваги та недоліки при проектуванні гідротехнічних об'єктів. Точні та наближені методи розв'язання інженерних задач. Похибки наближених розв'язків. Коректність задач, стійкість та збіжність наближених методів. Поняття про наближення функції. Точкова та неперервна апроксимація. Лінійна, кусочно-лінійна та квадратична інтерполяція. Алгоритм як основа математичного методу. Ітераційні методи. Метод скінченних різниць. Кінцево-різницева схема. Задачі інтер- та екстраполяції. Використання засобів *MS Excel* для розв'язування задач. Макроси та їх застосування для автоматизації роботи з даними в *MS Excel*.

Тема 8. Методи розв'язання систем лінійних алгебраїчних рівнянь (СЛАР)

Інженерні задачі при проектуванні гідротехнічних об'єктів, що приводять до СЛАР. Методи аналітичного та чисельного розв'язання СЛАР. Методи Гауса, Якобі, Зейделя. Методи, пов'язані зі застосуванням матричних процедур та дій з визначниками.

Тема 9. Чисельне розв'язання прикладних інженерних задач на основі диференціальних рівнянь

Приклади математичних моделей на основі диференціальних рівнянь в частинних похідних в гідродинаміці. Рівняння руху ідеальної рідини. Рівняння руху нестисненої в’язкої рідини. Рівняння неусталеного руху води у відкритих руслах. Чисельне диференціювання та інтегрування функцій. Апроксимація похідних функції кінцевими різницями. Похибка апроксимації та способи підвищення точності. Звичайні диференціальні рівняння першого порядку, задача Коші та крайова задача. Методи Ейлера та Рунге-Кутта. Обчислення визначених інтегралів. Метод Сімпсона.

Тема 10.Чисельне розв’язання задачі оптимізації

Невизначеність факторів. Основні поняття і загальна схема вирішення задач оптимізації. Критерій оптимізації і цільова функція. Математична постановка задачі оптимізації. Методи одновимірної оптимізації: градієнтного спуску, золотого перерізу, половинного оберненого кроку. Поняття про багатовимірну оптимізацію. Задача лінійного та нелінійного програмування. Розв’язання задач оптимізації в середовищі *MS Excel*. Засоби аналізу даних *Підбір параметру* і *Пошук рішення*.

1.2. Перелік тем практичних занять

1. Статистична обробка вибіркових даних. Оцінка результату прямих вимірювань за вибірковими даними.
2. Середньоквадратичне відхилення непрямих вимірювань.
3. Перевірка вибіркових даних на однорідність.
4. Побудова емпіричних функцій розподілу та забезпеченості вибіркових даних.
5. Кореляційний аналіз рядів даних.
6. Аналіз рядів даних за допомогою функцій середовища програмування R program.
7. Визначення коефіцієнтів регресії.
8. Перевірка адекватності рівняння регресії.
9. Розв’язання систем лінійних алгебраїчних рівнянь ітераційними методами.
10. Чисельне розв’язання задачі Коші для звичайного диференціального рівняння.
11. Чисельне інтегрування.
12. Задача лінійного програмування.
13. Задача нелінійного програмування.

1.3. Перелік тем до самостійної роботи студента

1. Приклади моделювання в інженерних задачах.
2. Різновиди математичних моделей. Аналітичне, чисельне, аналогове, імітаційне моделювання.
3. Види аналітичних законів розподілу випадкових величин.
4. Кореляція і причинно-наслідковий зв'язок.
5. Значення даних спостережень в проектуванні та експлуатації гідротехнічних об'єктів.
6. Прогнозування на основі регресійних моделей. Репрезентативність моделі.
7. Збіжність ітераційних процесів.
8. Точність методів Гауса, Якобі, Зейделя при розв'язанні СЛАР.
9. Оцінка похибки чисельного інтегрування.
10. Графічне розв'язання задач оптимізації.

2. Завдання до практичних занять та самостійної роботи студентів

2.1. Статистичні параметри вибірки

Випадкова величина – та, про яку немає повної інформації, але відомі деякі значення, які вона приймає. При проведенні експерименту (випробування, спостереження) випадкова величина може приймати різні значення з певною ймовірністю.

Вибірка або ***вибіркова сукупність*** значень випадкової величини – репрезентативна частина ***генеральної сукупності*** (всієї множини можливих значень випадкової величини), яка відбирається для безпосереднього вивчення величини або залучення до дослідження.

Описова статистика — набір статистичних показників, що характеризують вибірку даних та можуть слугувати для представлення всієї сукупності. Основними параметрами вибірки є:

– ***математичне сподівання*** (центр розподілу, для дискретних числових випадкових величин – ***вибіркове середнє***) є оцінкою (описує) випадкову величину за її абсолютним значенням

$$x = \frac{\sum_{i=1}^{n} x_i}{n}; \qquad (2.\ 1)$$

– ***дисперсія*** характеризує ступінь розсіяння елементів сукупності навколо середнього (є оцінкою точності)

$$D(x) = S^2(x) = \frac{\sum_{i=1}^{n}\left(x_i - x\right)^2}{n-1}; \qquad (2.\ 2)$$

– ***середньоквадратичне відхилення (СКВ, стандартне відхилення, середньоквадратична похибка, стандарт)*** характеризує стандартне відхилення вибіркового середнього

$$S(x) = \sqrt{\left(\sum_{i=1}^{n}\left(x_i - x\right)^2\right)/\left(n-1\right)}; \qquad (2.\ 3)$$

– ***коефіцієнт варіації*** характеризує мінливість, ступінь коливання досліджуваної величини

$$C_v(x) = \frac{\sigma(x)}{\bar{x}}; \qquad (2.\ 4)$$

– ***коефіцієнт асиметрії*** характеризує силу та напрямок асиметрії вибіркового розподілу (зміщення більшості членів ряду відносно середнього)

$$C_s(x) = \frac{A(x)}{\sigma(x)^3}; \qquad (2.\ 5)$$

– ***вибіркова асиметрія***

$$A(x) = \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)(n-2)}. \qquad (2.\ 6)$$

Під ***надійністю*** або ***довірчою ймовірністю*** випадкової величини (результату експерименту, спостережень) розуміють ймовірність того, що істинне значення величини, що досліджується, знаходиться в межах ***довірчого інтервалу (інтервалу довіри)***, границі якого визначають як $x \pm \Delta x$.

При перевірці ***статистичних гіпотез*** (припущень про властивості рядів даних) використовують певні модельні розподіли, які служать для виявлення міри розбіжності між емпіричними і гіпотетичними значеннями. ***Статистичний критерій*** — строге математичне правило, за яким приймається або відкидається статистична гіпотеза.

Завдання 1. При повторному вимірюванні величини витрати води при проведенні експериментальних досліджень руху води в каналі в лабораторних умовах отримано ряд (вибірку) даних. Виконати обробку ряду даних, виключити грубі помилки та записати результат. При перевірці статистичних гіпотез рівень значимості прийняти 0,05. Вихідні дані наведено в табл.2.1.

Таблиця 2. 1

№ вар.	Q_i, см^3/с									
1	0,51	0,58	0,56	0,55	0,56	0,57	0,54	0,53	0,54	0,57
2	2,4	2,5	2,4	2,6	2,3	2,7	2,3	2,5	2,2	2,4
3	126	124	128	126	121	125	127	123	126	124
4	6,1	6,4	6,5	6,3	6,6	6,8	6,4	6,9	6,7	6,6

Порядок виконання:

1. Визначити оцінки параметрів вибірки:

– математичного сподівання $\bar{x}$ – за (2.1) або із використанням вбудованої функції *MS Excel СРЗНАЧ(* x_i) (*AVERAGE*);

– СКВ $S(x)$ за (2.3) або із використанням вбудованої функції *MS Excel СТАНДОТКЛОН(* x_i) (*ST.DEV*);

2. Виключити грубі помилки та (за потреби) повторно визначити математичне сподівання та оцінку СКВ:

– ранжувати елементи вибірки в порядку зростання;

– визначити максимальне відносне відхилення для крайніх елементів вибірки (максимального та мінімального значень)

$$v = \frac{|\Delta x_{\max}|}{S(x)}; \qquad (2.\ 7)$$

– перевірити гіпотезу про помилковість елементу. Результат є грубо помилковим при виконанні умови

$$v > v_{1-p}, \qquad (2.\ 8)$$

де v_{1-p} – квантиль розподілу максимального відносного відхилення (дод. 1);

– при підтвердженні гіпотези про помилковість результату він виключається із ряду даних та проводиться повторне визначення параметрів вибірки (п. **1**);

– провести повторну перевірку вибірки на наявність грубих помилок.

6. Визначити напівширину довірчого інтервалу за відомим законом розподілу випадкової величини (визначення відхилення генерального середнього). Припускається, що закон розподілу нормальний.

$$\Delta x = S(x) \cdot t = \frac{S(x)}{\sqrt{n}} \cdot t\,, \qquad (2.\ 9)$$

де t_p– квантиль розподілу Стьюдента довірчої ймовірності p (дод. 2).

7. Визначити границі довірчого інтервалу. Сформувати результат вимірювань – оцінку генерального середнього відповідно до виразу

$$a = x \pm \Delta x\,. \qquad (2.\ 10)$$

8. Перевірити правильність визначення параметрів вибірки за допомогою вбудованого ***Сервісу “Аналіз даних”*** *MS Excel*. Для цього необхідно активувати пакет аналізу через вкладку **Файл** пункт **Параметри/ Надбудови/ Перейти,** відмітити необхідні для інсталювання надбудови. Доступ до пакетів аналізу – через вкладку **Дані/ Аналіз даних/ Описова статистика**.

Приклад розв’язання завдання 1.

Результати вимірювань $Q_i; i \in [1...8]$ представлено в табл. 2.2.

Таблиця 2. 2

i	1	2	3	4	5	6	7	8
Q_i, см³/с	485	482	485	483	488	483	486	484

1. Після ранжування ряду визначено крайні елементи: 488 та 482.

2. Перевірка на помилковість крайнього елементу вибірки Q_8=488 см³/с:

– визначення параметрів вибірки без врахування сумнівного результату:

$\overline{Q} = СРЗНАЧ\left(Q_1 : Q_7\right) = 484$ (см³/с);

S(Q)=СТАНДОТКЛОН(Q₁:Q₇)=1,41 (см³/с).

– відносне відхилення $v = \frac{|488 - 484|}{1{,}41} = 2{,}828$;

– критичне значення v_{1-p} для p=0,95 та n=7 $v_{0,05}$=2,093;

– перевірка умови (2.8): 2,828>2,093 – умова виконується, отже, елемент вибірки Q_8=488 см^3/с є помилковим; його виключають із ряду даних.

3. Перевірка на помилковість крайнього елементу Q_1=482см^3/с:

– визначення параметрів вибірки без врахування сумнівного результату (ряду із шести значень):

$\overline{Q} = СРЗНАЧ\left(Q_2 : Q_7\right) = 484{,}3$ (см^3/с);

S(Q)=СТАНДОТКЛОН(Q₂:Q₇)=1,21 (см^3/с).

– відносне відхилення $\nu = \dfrac{|482 - 484{,}3|}{1{,}21} = 1{,}927$;

– табличне значення ν_{1-p} для p=0,95 та n=6 $\nu_{0,05}$=1,996;

– перевірка умови (2.8): 1,927<1,996 – умова не виконується, отже, результат Q_1=482 см^3/с не є помилковим.

4. Перевірка на помилковість крайнього елементу Q_7=486см^3/с:

$\overline{Q} = СРЗНАЧ\left(Q_1 : Q_6\right) = 483{,}7$ (см^3/с);

S(Q)=СТАНДОТКЛОН(Q₁:Q₆)=1,21 (см^3/с);

– відносне відхилення $\nu = \dfrac{|486 - 483{,}7|}{1{,}21} = 1{,}927$;

– табличне значення ν_{1-p} для p=0,95 та n=6 $\nu_{0,05}$=1,996;

– перевірка умови (2.8): 1,927<1,996 – умова не виконується, отже, результат Q_7=486 см^3/с не є помилковим.

5. Параметри виправленого ряду даних (із шести значень):

$\overline{Q} = СРЗНАЧ\left(Q_1 : Q_7\right) = 484$ (см^3/с);

S(Q)=СТАНДОТКЛОН(Q₁:Q₇)=1,41 (см^3/с).

6. Визначення напівширини довірчого інтервалу за (2.9)

$$\Delta\overline{Q} = S(\overline{Q}) \cdot t = \frac{S(Q)}{\sqrt{n}} \cdot t = \frac{1{,}41}{\sqrt{7}} \cdot 2{,}447 = 1{,}3 \text{ (см}^3\text{/с)}.$$

<u>Відповідь:</u> Результат вимірювання із надійністю p=0,95 знаходиться в інтервалі Q=484±1,3 см^3/с. Границі довірчого інтервалу: 482,7 см^3/с ≤ Q ≤485,3 см^3/с.

	A	B
1	*Row1*	
2		
3	Mean	484.5
4	Standard E	0.681385
5	Median	484.5
6	Mode	485
7	Standard [	1.927248
8	Sample Va	3.714286
9	Kurtosis	0.182249
10	Skewness	0.638615
11	Range	6
12	Minimum	482
13	Maximum	488
14	Sum	3876
15	Count	8

Рис. 2. 1. Визначення параметрів вибірки за допомогою вбудованого ***Сервісу “Аналіз даних”*** *MS Excel*.

2.2. Визначення середньоквадратичного відхилення результату непрямих вимірювань

Якщо деяка величина z залежить від декількох незалежних величин x_i $z = f(x_1, x_2, ..., x_n)$, СКВ яких $S(x_i)$ відомі, середньоквадратичне відхилення цієї величини визначають за залежністю

$$S(z) = \sqrt{\left(\frac{\partial z}{\partial x_1} S(x_1)\right)^2 + \left(\frac{\partial z}{\partial x_2} S(x_2)\right)^2 + ... + \left(\frac{\partial z}{\partial x_n} S(x_n)\right)^2}, \qquad (2.\ 11)$$

де $\frac{\partial z}{\partial x_i}$ – часткові похідні функції $z = f(x_1, x_2, ..., x_n)$.

В частковому випадку, якщо величина z залежить від однієї змінної z=$f(x)$, формула (2.11) спрощується до виду

$$S(z) = \frac{dz}{dx} S(x). \qquad (2.\ 12)$$

Завдання 2. Визначити середньоквадратичне відхилення результату непрямих вимірювань, якщо задані СКВ величин, отриманих прямими вимірюваннями, та вид залежності між величинами. Вихідні дані прийняти за табл. 2.3.

Таблиця 2. 3

№ вар.	Вид залежності	Величина, отримана в результаті прямих вимірювань		
		x_1	x_2	x_3
1	$J = (H_1 - H_2)/l$	H_1=50см $S(H_1)$=0,1см	$H2$=35 см $S(H_2)$=0,1 см	l=75 см $S(l)$=0,05 см
2	$Q = W/t$	W=40 л $S(W)$=10 мл	t=352 с $S(t)$=0,2 с	
3	$\mu = \dfrac{Q}{\frac{\pi d^2}{4}\sqrt{2g \cdot z}}$	Q=0,5 см^3/с $S(Q)$=0,005 см^3/с	d=0,4 см $S(d)$=0,01 см	z=6,1 см $S(z)$=0,1 см
4	$V = \dfrac{Q}{b \cdot h}$	Q=3,6 см^3/с $S(Q)$=0,1 см^3/с	b=2,5 см $S(b)$=0,1 см	h=8,2 см $S(h)$=0,1 см
5	$\mu = \dfrac{V}{\sqrt{2g \cdot z}}$	V=3,6 см/с $S(V)$= 0,1 см/с	z=2,5 см $S(z)$=0,1 см	
6	$Q = \mu\omega\sqrt{2g \cdot z}$	ω=4,5 см^2 $S(\omega)$=0,01 см^2	z=3,8 см $S(z)$=0,1 см	μ=0,6

Приклад розв’язання Завдання 2. Знайти середньоквадратичне відхилення величини витрати *Q*, виміряної за допомогою трикутного мірного водозливу, якщо СКВ при вимірюванні напору *H* на гребені водозливу *S(H)*=0,001 м. Значення напору *H*=0,1м. Вид залежності між витратою та напором: $Q = 1{,}4H^{5/2}$.

Припускаємо, що мірна голка водозливу тарована відносно нуля водозливу, тому *Q* є функцією однієї змінної *Q =f(H)*.

$$S(Q) = \frac{dQ}{dH} S(H);$$

$$\frac{dQ}{dH} = \frac{d\left(1{,}4H^{5/2}\right)}{dH} = \frac{5}{2} \cdot 1{,}4 \cdot H^{\frac{3}{2}};$$

За (2.12): $S(Q) = 3{,}5 \cdot 0{,}1^{3/2} \cdot 0{,}001 = 1{,}1 \cdot 10^{-4} м^3/с = 0{,}11 л/с$.

Відповідь: *S(Q)*=0,11 л/с.

2.3. Побудова емпіричної функцій розподілу за вибірковими даними

Для опису випадкових величин використовують ***закон розподілу ймовірностей*** – функцію, яка для випадкової величини показує множину можливих її значень із ймовірністю (або частотою) їх появи.

За наявними емпіричними даними будують емпіричну функцію розподілу. Теоретичною функцією розподілу випадкової величини *x* називають функцію дійсного аргументу, що задається у вигляді $F(x)=P(X \le x)$.

Для побудови ***емпіричного розподілу*** $F_{m(n)}$ виконують ранжування членів вибірки в зростаючому порядку; для побудови емпіричної функції забезпеченості $G_{m(n)}$ – в спадаючому порядку. Емпіричну ймовірність $F_{m(n)} = P_{m(n)}$ або емпіричну забезпеченість $G_{m(n)} = P_{m(n)}$ i-го значення випадкової величини можна визначити за формулою Чегодаєва

$$P_{m(n)} = \frac{m - 0{,}3}{n + 0{,}4}, \qquad (2.\ 13)$$

де m – порядковий номер ранжируваного в зростаючому (при визначенні $F_{m(n)}$) чи спадаючому (при визначенні $G_{m(n)}$) порядку ряду значень, n – кількість значень у вибірці.

Завдання 3. Побудувати інтегральну криву емпіричного розподілу вибіркових даних, наведених в табл. 2.4. При розрахунку емпіричної ймовірності використати формулу Чегодаєва.

Таблиця 2. 4

Вибіркові дані x_i

i	x_i				i	x_i			
	номер варіанту					номер варіанту			
	1	2	3	4		1	2	3	4
1	10400	7750	6700	2710	26	4840	5520	4930	4890
2	4900	4530	5120	3410	27	6740	7280	5670	2120
3	11050	5500	1930	7660	28	5240	3430	3340	9820
4	2040	2240	2500	7530	29	1660	2330	2010	6820
5	3030	2530	1780	10840	30	5440	2940	7560	9570
6	5650	3570	7000	3420	31	2470	2500	3720	1670
7	3030	1720	12270	4180	32	3750	5650	3590	3730
8	2020	2200	7820	3770	33	4290	9170	2860	4290
9	6070	3850	5220	1570	34	10240	3630	2520	3020
10	3030	5000	3870	7200	35	4680	1330	5010	5590
11	10380	4360	5050	2900	36	3110	1520	4170	3000
12	3030	5280	2580	9790	37	2390	5960	8380	2930
13	5400	7820	6720	1090	38	2240	5220	8660	1610
14	8830	1420	2850	6060	39	5280	6400	5030	5820
15	5030	1600	2750	4310	40	2700	6800	2500	1950
16	3730	5860	5330	4910	41	3120	2100	3890	5560
17	5240	13600	3040	7740	42	3900	3280	5310	1790
18	4640	2040	8760	1620	43	3920	5540	2840	3000
19	5650	5030	7530	15300	44	11090	4700	3660	2370
20	3570	3700	3920	9140	45	2830	4250	9150	2200
21	3730	11190	5100	3640	46	11140	4650	3330	3200
22	9030	8560	2930	7950	47	2870	12300	4310	2300
23	5030	3850	4540	6530	48	2240	4600	3330	3200
24	2180	3280	2180	3730	49	2270	2160	3740	4800
25	2470	5340	7620	4160	50	6140	2480	1680	2100

Порядок виконання:

1. Ранжувати члени вибірки в зростаючому порядку та присвоїти їм порядковий номер.

2. Ординати інтегральної кривої емпіричного розподілу розрахувати за формулою (2.13).

3. Створити діаграму, розташувавши значення вибірки по осі X та розраховані кумулятивні ймовірності по осі Y.

2.4. Перевірка вибіркових даних на відповідність нормальному закону розподілу

Найбільш широко використовуваним аналітичним законом розподілу в статистиці є ***нормальний розподіл***. Його використовують, коли випадкова величина залежить від великої кількості випадкових факторів, однорідних за своїм впливом, а також, коли вибірки мають значний об'єм.

Функція ***щільності нормального розподілу***

$$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}}\exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x{}^2}\right), \qquad (2.14)$$

інтегральна функція нормального розподілу

$$F(x) = \frac{1}{\sigma_x\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x{}^2}\right). \qquad (2.15)$$

Ординати ***функції забезпеченості*** при відомій ординаті інтегральної функції розподілу можна розрахувати за залежністю

$$G(x) = 1 - F(x). \qquad (2.16)$$

Оскільки стандартні методи розрахунку статистичних параметрів, як правило, розроблені, виходячи із припущення, що розподіл випадкової величини є нормальним, при обробці вибіркових даних для отримання коректних результатів необхідне проведення перевірки гіпотези відповідності вибірки нормальному розподілу.

Одним із статистичних критеріїв для перевірки емпіричного закону розподілу на нормальність є ***критерій Колмогорова–Смирнова***, що дозволяє порівнювати два розподіли за кількісним показником $D_{n,m}$, наприклад, визначає відстань між емпіричною функцією розподілу вибірки та кумулятивною функцією еталонного розподілу.

$$\boldsymbol{D_{n,m} = \sup_{x}\left|F_{1,n}(x) - F_{2,m}(x)\right|}, \qquad (2.17)$$

де $F_{1,n}(x); F_{2,m}(x)$ – ординати функцій розподілу, що порівнюють;

sup – верхня межа (максимальне значення) для підмножини в частково впорядкованій множині.

При виконанні умови

$$\sqrt{n}D_{n,m} > K_{n,\alpha} \qquad (2.\ 18)$$

нуль-гіпотеза про однорідність вибірок (у випадку порівняння емпіричного розподілу із нормальним – гіпотеза про нормальний закон розподілу вибірки) відкидається; де $K_{n,\alpha}$ – квантиль розподілу Колмогорова для заданого рівня значимості α (дод. 3).

Завдання 4. Провести перевірку емпіричного розподілу із завдання 3 на нормальність. Для перевірки використати критерій Колмогорова–Смирнова.

Порядок виконання:

1. Визначити математичне сподівання та середньоквадратичне відхилення вибірки за формулами (2.1) та (2.3) або із використанням вбудованих функцій *MS Excel*.

2. Визначити ординати інтегральної кривої нормального розподілу за формулою (2.15) або за допомогою вбудованої функцій *MS Excel НОРМРАСП(x,* $\bar{x}$, $S(x)$, *TRUE)* (*NORM.DIST*) за значеннями математичного сподівання та СКВ вибірки та логічного аргументу (для введення значення аргументу *TRUE* ввести 1).

3. Визначити абсолютне значення різниць ординат емпіричного та нормального законів розподілу.

4. Перевірити виконання умови (2.18) та зробити висновок про прийняття чи відхилення гіпотези про нормальний розподіл вибірки.

Приклад розв'язання завдання 4 наведено в табл. 2.5. Ранжовані вибіркові дані представлено в стовбці 2 таблиці.

Таблиця 2. 5

№ з/п	x_i	$F_{1,n}(x) = P_{m(n)}$ за (2.13)	$F_{2,m}(x)$ за (2.15)	$F_{1,n}(x) - F_{2,m}(x)$
1	2	3	4	5
1	1,2	0,067	0,140	0,073
2	1,6	0,133	0,190	0,057
3	1,7	0,200	0,204	0,004
4	1,8	0,267	0,219	0,048
5	1,9	0,333	0,234	0,100
6	2	0,400	0,249	0,151

№ з/п	x_i	$F_{1,n}(x) = P_{m(n)}$ за (2.13)	$F_{2,m}(x)$ за (2.15)	$F_{1,n}(x) - F_{2,m}(x)$
1	2	3	4	5
7	2,2	0,467	0,283	0,184
8	2,6	0,533	0,354	0,179
9	3	0,600	0,432	0,168
10	3,5	0,667	0,532	0,134
11	4	0,733	0,630	0,103
12	4,8	0,800	0,769	0,031
13	5,6	0,867	0,873	0,006
14	6,6	0,933	0,950	0,017
15	7,6	1	0,984	0,016

$x =$ 3,34 $D_{n,m} =$ 0,184

$S(x) =$ 1,982 $\frac{\boldsymbol{K}_{15,0.05}}{\sqrt{15}} =$ 0,338

Перевірка умови (2.18): 0,184<0,338 – умова не виконується, отже, для опису вибірки може бути використана модель нормального розподілу.

2.5. Обробка результатів груп (серій) вимірювань

Групи (серії) вимірювань можуть бути отримані, наприклад, при проведенні повторних вимірювань деякої величини на різних рівнях (при різних значеннях) вхідного фактора (незалежної змінної). При обробці серій вимірювань необхідним є перевірка гіпотези про однорідність дисперсій вибірок.

Однорідність (незначимість варіації) дисперсій означає, що дисперсії вимірювань на різних рівнях практично однакові (однорідні), тобто точність вимірювань може вважатися однаковою для всіх серій досліджень. Точність вимірювань при проведенні експерименту в такому випадку характеризує значення середньої дисперсії.

При однаковому розмірі вибірок, що порівнюють, однорідність дисперсій перевіряють за G – критерієм Кохрана

$$G = \frac{S^2(x)_{\max}}{\sum_{i=1}^{k-1} S^2(x)_i}, \qquad (2.\ 19)$$

де $S^2(x)_{\max}$ – максимальна дисперсія; $\sum_{i=1}^{k-1} S^2(x)_i$ – сума решти дисперсій; k – кількість дисперсій, що порівнюють.

Нуль-гіпотеза про незначимість варіації (рівність, однорідність) дисперсій приймається, якщо отримане значення критерію Кохрана задовольняє умову

$$G < G_{1-p}, \tag{2. 20}$$

де $G_{1-p}=\varphi(p;k;f)$ – квантиль розподілу Кохрана (дод. 4); f - кількість ступенів вільності дисперсій; $f=n$-1; n – кількість елементів вибірки (повторних вимірювань в серії); 1-p – рівень значимості.

Завдання 5. В результаті вимірювань придонних швидкостей потоку гідрометричною вертушкою при різних витратах отримано серії вимірювань. Виконати перевірку однорідності дисперсій серій вимірювань. Визначити оцінку середньої дисперсії. Вихідні дані наведено в табл. 2.6.

Таблиця 2. 6

№з/п	V	h_1	h_2	h_3	V	h_1	h_2	h_3
	Варіант 1				Варіант 2			
1	0	53,1	54,5	53,5	0	5,13	5,21	5,28
2	1	42,8	43,6	42,6	5	6,25	6,35	6,26
3	2	35,5	36,1	35,8	10	7,32	7,45	7,35
4	3	26,9	25,8	26,1	15	8,52	8,48	8,4
5	4	14,7	15,2	15,8	20	9,84	9,75	9,88
	Варіант 3				Варіант 4			
1	0	0,164	0,168	0,177	20	133	135	136
2	2	0,347	0,358	0,349	30	115	113	116
3	4	0,565	0,558	0,59	40	93	95	97
4	6	0,726	0,736	0,738	50	75	78	74
5	8	0,923	0,925	0,938	60	56	58	59

Послідовність виконання:

1. Для кожної із вибірок визначити оцінки параметрів закону розподілу: математичного сподівання за формулою (2.1) та дисперсії за формулою (2.2).

2. Обчислити значення критерію Кохрана за формулою (2.19).

3. Перевірити гіпотезу про однорідність дисперсій вибірок за умовою (2.20).

4. Якщо гіпотезу про однорідність дисперсій підтверджено, визначити оцінку середньої дисперсії

$$\bar{S}^2(x) = \frac{\sum_{i=1}^{k} S^2(x)_i}{k}. \qquad (2.\ 21)$$

<u>Приклад розв'язання завдання 5.</u>

Результати експериментальних досліджень представлено в стовбцях 2–6 табл. 2.7. Математичне сподівання та дисперсію для кожної із вибірок наведено в стовбцях 7 та 8 відповідно.

Таблиця 2. 7

№з/п	x	y_1	y_2	y_3	y_4	$\bar{y}$	$S^2(y)$
1	2	3	4	5	6	7	8
1	0	55,3	54,6	55,1	55,2	55,1	0,0967
2	10	63,8	64,2	63,9	64,1	64,0	0,0333
3	20	67,2	67,1	66,9	66,8	67,0	0,0333
4	30	74,3	73,8	74,1	74,2	74,1	0,0467
5	40	81,9	82,2	82,3	81,8	82,1	0,0567
						$\sum_{i=1}^{k-1} S^2(y)_i$	0,1700
						$\sum_{i=1}^{k} S^2(y)_i$	0,2667

Значення критерію Кохрана $G = \frac{0,0967}{0,1700} = 0,5686$. Критичне значення критерію Кохрана при кількості ступенів вільності f=3; k=5 $G_{0,05}$=0,5981.

Перевірка умови (2.20): 0,5686<0,5981 – умова виконується, отже, гіпотеза про однорідність дисперсій приймається.

Середня дисперсія $\bar{S}^2(y) = 0,2667/5 = 0,0533$.

<u>Відповідь:</u> Дисперсії груп вимірювань можна вважати однорідними. Оцінка середньої дисперсії для всієї серії даних $\bar{S}^2(y) = 0,0533$.

2.6. Кореляційний аналіз

Функціональний зв'язок між двома величинами має місце, якщо кожному значенню незалежної змінної (x) відповідає певне значення залежної змінної (y).

При ***статистичному (стохастичному) зв'язку*** одному і тому ж значенню однієї змінної можуть відповідати різні значення іншої. Наявність статистичного зв'язку пояснюється тим, що величини, що досліджуються, змінюються не тільки в силу взаємодії між собою, але й під впливом багатьох випадкових факторів,

Кореляція – залежність між випадковими величинами, що не має строго функціонального характеру; натомість має місце статистичний, стохастичний зв'язок.

Кореляційний аналіз – встановлення наявності та тісноти зв'язку між величинами та ступеня значимості (невипадковості) зміни випадкової величини в процесі досліджень. При кореляційному аналізі перевіряють гіпотезу про значимість коефіцієнта кореляції.

Коефіцієнт кореляції – критерій оцінки стохастичного зв'язку між величинами, міра кореляційного зв'язку. Значення коефіцієнта кореляції знаходиться в межах від -1,0 до +1,0.

Для визначення ***коефіцієнта парної кореляції (Пірсона)*** використовують залежність

$$r = \frac{\sum_{i=1}^{n} x_i \cdot y_i - n \cdot x \cdot y}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n \cdot (x)^2\right)\left(\sum_{i=1}^{n} y_i^2 - n \cdot (y)^2\right)}}, \qquad (2.\ 22)$$

або

$$r = \frac{\sum_{i=1}^{n} (x_i - x) \cdot (y_i - y)}{S(x) \cdot S(y) \cdot (n-1)}. \qquad (2.\ 23)$$

При оцінці тісноти кореляційного зв'язку можна використати такі орієнтовні інтервали значень. Дві випадкові величини:

– незалежні, дуже слабка кореляція – при значенні коефіцієнта кореляції в межах $0 < |r| \le 0{,}2$;

– слабокорельовані – $0{,}2 < |r| \le 0{,}5$;

– середньокорельовані – $0{,}5 < |r| \le 0{,}7$;

– сильнокорельовані $0{,}7 < |r| \le 0{,}9$;

– наявна лінійна залежність $0{,}9 < |r| \le 1$.

Завдання 6. Встановити наявність та тісноту кореляційного зв'язку між двома випадковими величинами. Використати вихідні дані із Завдання 5. Рівень значимості прийняти 10%.

Послідовність виконання:

1. Визначити коефіцієнт кореляції за формулами (2.22) або (2.23).

2. Величини є корельованими (зміна випадкових величин статистично значима), якщо виконується умова

$$r > r_{1-p/2}, \qquad (2.\ 24)$$

де $r_{1\text{-}p/2}$ – квантиль розподілу вибіркового коефіцієнта кореляції (дод. 5); $r_{1\text{-}p/2}(f)$; $f=n\text{-}1$ – кількість ступенів вільності вибірки об'ємом n.

3. Перевірити правильність обчислень за допомогою вбудованої функції *MS Excel* $r = КОРРЕЛ\,(x_1 : x_n; y_1 : y_n)$ (*CORREL*).

Приклад розв'язання завдання 6.

Оброблені дані експериментальних досліджень прийнято за стовбцями 1, 7 табл. 2.7. Розрахунки для проведення кореляційного аналізу представлено у вигляді табл. 2.8.

Таблиця 2. 8

№з/п	x	y	x^2	$x \cdot y$	y^2	$x_i - x$	$(x_i - x)^2$	$y_i - y$	$(y_i - y)^2$	$(x_i - x) \times (y_i - y)$
1	2	3	4	5	6	7	8	9	10	11
1	0	55,1	0	0	3036,01	-20	400	-13,4	178,49	267,2
2	10	64	100	640	4096	-10	100	-4,5	19,89	44,6
3	20	67	400	1340	4489	0	0	-1,5	2,13	0
4	30	74,1	900	2223	5490,81	10	100	5,6	31,81	56,4
5	40	82,1	1600	3284	6740,41	20	400	13,6	186,05	272,8
$\sum$	100	342,3	3000	7487	23852,2		1000		418,37	641
Сер.	20	68,5								

За формулою (2.22):

$$r = \frac{7487 - 5 \cdot 20 \cdot 68{,}5}{\sqrt{\left(3000 - 5 \cdot 20^2\right) \cdot \left(23852{,}23 - 5 \cdot 68{,}5^2\right)}} = 0{,}991;$$

або за формулою (2.23): $S(x) = \sqrt{\dfrac{1000}{5-1}} = 15{,}8;$

$$S(y) = \sqrt{\frac{418{,}37}{5-1}} = 10{,}23;$$

$$r = \frac{641}{15{,}8 \cdot 10{,}23 \cdot (5-1)} = 0{,}991 \cdot$$

Критичне значення коефіцієнта кореляції при кількості ступенів вільності f=4 $r_{1-p/2}$=0,7293.

Перевірка умови (2.24): 0,991>0,7293 – умова виконується, отже, гіпотеза про наявність статистично значимого кореляційного зв'язку між величинами на рівні 0,1 приймається.

Відповідь: Величини корельовані (взаємозалежні),

2.7. Визначення коефіцієнтів лінійної регресії методом найменших квадратів

Математична модель – система математичних співвідношень: рівнянь, нерівностей, – що відображають істотні властивості об'єкту чи явища.

Регресія – математична модель, що використовується в математичній статистиці для опису залежності між випадковими величинами. Регресійна модель повинна відтворювати характер експериментальної залежності (тобто бути максимально близькою до експериментальних точок), але в той самий час згладжувати випадкові відхилення функції відгуку.

Регресійний аналіз – встановлення характеру зв'язку між величинами (виду регресії: лінійна, експоненціальна, періодична тощо) та визначення коефіцієнтів ***рівняння регресії*** .

Метод найменших квадратів (МНК) (принцип максимальної правдоподібності) – метод пошуку коефіцієнтів рівняння регресії, при значеннях яких сума квадратів відхилень значень емпіричної функції та рівняння регресії є мінімальною.

$$F(b_0,b_1)=\sum_{i=1}^{N}\left(y_i-f(x_i)\right)^2=\sum_{i=1}^{N}\left(y_i-(b_0+b_1x_i)\right)^2\to\min. \quad (2.25)$$

Для лінійної регресії $f(x_i)=b_0+b_1\cdot x$ умова (2.25) виконується, якщо

$$\begin{cases}\dfrac{\partial F(b_0,b_1)}{\partial b_0}=0\\[2mm]\dfrac{\partial F(b_0,b_1)}{\partial b_1}=0\end{cases}. \quad (2.26)$$

Розв'язки системи (2.26) для лінійної регресії:

$$b_1=\frac{N\cdot\sum_{i=1}^{N}x_i\cdot y_i-\sum_{i=1}^{N}x_i\cdot\sum_{i=1}^{N}y_i}{N\sum_{i=1}^{N}x_i^2-\left(\sum_{i=1}^{N}x_i\right)^2}; \quad (2.27)$$

$$b_0 = \frac{\sum_{i=1}^{N} x_i^2 \cdot \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i \cdot y_i}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2}. \qquad (2.\ 28)$$

Завдання 7. Визначити коефіцієнти лінійної регресії для експериментальних даних із Завдання 5.

Послідовність виконання:

1. Визначити коефіцієнти лінійної регресії за (2.27) та (2.28).

2. Записати рівняння регресії у вигляді $\tilde{y} = b_0 + b_1 \cdot x$.

3. Перевірити правильність розрахунків, обчисливши параметри регресії за допомогою вбудованої функції *MS Excel ЛИНЕЙН (LINEST).* Аргументами функції є масиви даних та логічні значення:

– Логічний аргумент *Константа* вказує на наявність (значення 1) або відсутність (0) вільного члена в рівнянні регресії;

– при значенні *Статистика,* рівним 1, додаткова інформація виводиться, при значенні 0 виводяться тільки параметри рівняння.

4. Побудувати графіки експериментальної функції $y=f(x_i)$ та рівняння регресії, використовуючи *Майстер Діаграм,* Тип діаграми "*Точкова*", для лінії регресії – із згладжуючими лініями.

5. Нанести на графік лінію тренда, тип "*Лінійна*", вивести рівняння лінії тренду на діаграму. Порівняти отримане рівняння регресії із рівнянням лінії тренду.

Приклад розв'язання Завдання 7. Вихідними даними є масиви значень x та y із табл. 2.8 (стовбці 2, 3). Результати обчислень зведено в табл. 2.9.

Таблиця 2. 9

№з/п	x	y	x^2	$x \cdot y$	$\tilde{y}$
1	0	55,1	0	0	55,64
2	10	64,0	640	100	62,05
3	20	67,0	1340	400	68,46
4	30	74,1	2223	900	74,87
5	40	82,1	3284	1600	81,28
$\sum$	100	342,3	7487	3000	55,64

$$b_0 = \frac{7487 \cdot 342.3 - 100 \cdot 3000}{5 \cdot 7487 - 100^2} = 55.64;$$

$$b_1 = \frac{5 \cdot 3000 - 100 \cdot 342.3}{5 \cdot 7487 - 100^2} = 0.641;$$

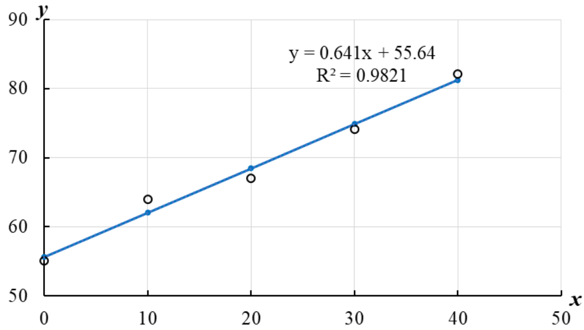


Рис. 2. 2. Графіки експериментальної функції та рівняння регресії

Відповідь: Рівняння лінійної регресії $\tilde{y} = 55.64 + 0.641 \cdot x$

2.8. Визначення коефіцієнтів нелінійної регресії

Застосовуючи МНК, відновлюють лише лінійні за параметрами функції. Але, використавши лінеаризуючі перетворення, цим методом можливо побудувати практично будь-які форми нелінійного парного зв'язку.

Лінеаризація моделі **–** представлення нелінійної залежності, що вивчається, у вигляді лінійного співвідношення між перетвореними змінними. Одним із методів лінеаризації є *логарифмування* із подальшою заміною змінних.

Приклад лінеаризації степеневої залежності виду $y = b_0 x^{b_1}$:

1. Логарифмування $\ln y = \ln b_0 + b_1 \cdot \ln x$; (2. 29)

2. Заміна змінних $Y = \ln y$; $B_0 = \ln b_0$; $X = \ln x$; (2. 30)

3. Отримання лінійної моделі $Y = B_0 + b_1 \cdot X$. (2. 31)

Приклад лінеаризації експоненціальної залежності виду $y = b_0 \cdot e^{b_1 \cdot x}$:

1. Логарифмування $\ln y = \ln b_0 + b_1 \cdot x$; (2. 32)

2. Заміна змінних – залежності (2.30);

3. Отримання лінійної моделі $Y = B_0 + b_1 \cdot x$. (2. 33)

Отримані лінеаризовані залежності можуть бути представлені графічно в логарифмічних (залежність (2.30) – (*X;Y*) або напівлогарифмічних (залежність (2.33) – (*x*; *Y*)) координатах.

Завдання 8. В результаті спостереження за динамікою відмов на водозливних греблях отриманий ряд даних (табл. 2.10). Підібрати коефіцієнти рівняння нелінійної регресії для експериментальних даних, застосувавши лінеаризацію.

Послідовність виконання:

1. Виконати логарифмування тих масивів даних, для яких це потрібно для заданого типу рівняння регресії.

2. Підібрати коефіцієнти лінійної регресії для лінеаризованої експериментальної залежності. Для цього скористатися раніше складеною програмою для МНК.

3. Записати рівняння нелінійної регресії для вихідних змінних.

4. Побудувати графіки експериментальної та регресійної залежностей для вихідних величин.

5. Нанести лінію тренду заданого типу, вивести рівняння лінії тренду на діаграму. Порівняти коефіцієнти рівняння регресії із рівнянням лінії тренду.

Таблиця 2. 10

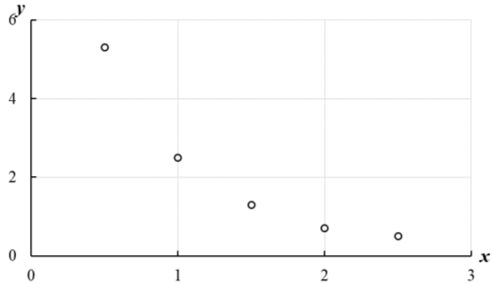
№ з/п	*t*, міс,	*p*, подій	*t*, міс,	*p*, подій
	Варіант 1		Варіант 2	
1	0,5	12,2	10	21,3
2	1	4	20	5,8
3	1,5	2,3	30	2,1
4	2	1,3	40	1,8
5	2,5	0,7	50	0,4
Тип регресії	степенева		степенева	
	Варіант 3		Варіант 4	
1	1	1,3	2	20,4
2	2	0,7	4	10,5
3	3	0,3	6	4,5
4	4	0,2	8	2,3
5	5	0,12	10	1,2
Тип регресії	експоненціальна		експоненціальна	

Приклад розв’язання Завдання 8.

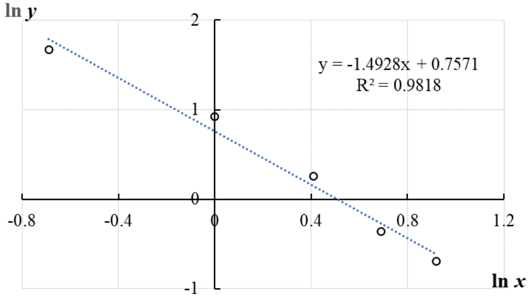
Вихідні дані наведено в стовбцях 2, 3 табл. 2.11.

Зроблено припущення про нелінійний зв’язок між величинами: степенева залежність виду $\tilde{y} = b_0 \cdot x^{b_1}$. Для лінеаризації вихідної залежності використано логарифмування.

Хід розрахунків зведено в табл. 2.11. Графік вихідної функції наведено на рис. 2.3 а), лінеаризованої залежності – на рис. 2.3 б).



а)



б)

Рис. 2. 3. Графіки експериментальної функції
а) вихідний; б) із перетвореними змінними

Таблиця 2. 11

№ з/п	x	y	$X = \ln x$	$Y = \ln y$
1	2	3	4	5
1	0,5	5,3	-0,69	1,67
2	1	2,5	0,00	0,92
3	1,5	1,3	0,41	0,26
4	2	0,7	0,69	-0,36
5	2,5	0,5	0,92	-0,69

Методом найменших квадратів до лінеаризованої залежності підібрано рівняння лінійної регресії виду $Y = B_0 + B_1 \cdot X$, в якому B_0=0,75; B_1= –1,49; (Y=0,75-1,49·X).

Для запису рівняння регресії у вихідних величинах виконано зворотні перетворення: $b_0 = e^{B_0}$ = $e^{0,75}$=2,12.

<u>Відповідь</u>: Рівняння степеневої регресії $\tilde{y} = 2{,}12 \cdot x^{-1{,}49}$.

2.9. Перевірка адекватності регресійної моделі

При відсутності *a priori* інформації про характер залежності між величинами в якості рівняння регресії використовують поліном степеня m

$$y = b_0 + b_1 \cdot x + b_2 \cdot x^2 + ... + b_m \cdot x^m . \qquad (2.\ 34)$$

Вибір оптимального виду регресії здійснюють методом послідовного її ускладнення (послідовних наближень), прийнявши в моделі (2.34) в першому наближенні m=1, в наступних – підвищуючи ступінь поліному.

Перевірку значимості (адекватності) рівняння регресії вибірковим даним можна проводити за ***F – критерієм Фішера***. Критерій Фішера дозволяє порівнювати будь-які дві дисперсії за залежністю

$$F = \frac{S_{a\partial}^2(y)}{S^2(y)}, \qquad (2.\ 35)$$

де $\bar{S}^2(y)$ – дисперсія відтворюваності

$$\bar{S}^2(y) = \frac{\sum_{i=1}^{N} S_i^2(y)}{N}; \qquad (2.\ 36)$$

$S_{a\partial}^2(y)$ – дисперсія адекватності

$$S_{a\partial}^2(y) = \frac{\sum_{i=1}^{N} \left(y_i - \tilde{y}_i\right)^2}{N-1}. \qquad (2.\ 37)$$

Рівняння регресії $\tilde{y} = f(x)$ адекватно описує експериментальні дані (x_i;y_i) при виконанні умови

$$F < F_{1-p}, \qquad (2.\ 38)$$

де $F_{1-p} = \varphi\left(f_1, f_2, p\right)$ – квантиль розподілу Фішера (дод. 6); f_1=N-m – кількість ступенів вільності при визначенні загальної дисперсії; f_2=N(n-1) – кількість ступенів вільності при визначенні залишкової дисперсії; n – кількість повторних вимірювань при визначенні дисперсії; p – рівень значимості.

Завдання 9. При лабораторному дослідженні розмивів нижнього б’єфу за гідротехнічними спорудами отримано ряди даних питомої витрати потоку q, см^3/с, глибини ями розмиву h_p, см та середнє значення дисперсії $\bar{S}^2(y)$ при вимірюванні h_p на кожному із рівнів (табл. 2.12). Кількість повторних вимірювань при проведенні дослідження n=3. Підібрати оптимальне рівняння регресії для експериментальних даних. Провести перевірку адекватності моделі за критерієм Фішера. Рівень значимості прийняти 0,05.

Таблиця 2. 12

№ з/п	q,см^3/с	h_p, см	$\bar{S}^2(y)$	q,см^3/с	h_p, см	$\bar{S}^2(y)$
	Варіант 1			Варіант 2		
1	1,7	25	1,5	2	4	1,5
2	3,4	44	1,3	4	13	1,3
3	3,8	57	1,6	6	27	1,6
4	4,1	65	1,5	8	48	1,5
5	5,3	98	1,4	10	72	1,4
6	6,3	141	1,3	12	113	1,3
7	7	175	1,6	14	152	1,6
	Варіант 3			Варіант 4		
1	0,5	34	1,5	1	6	1,5
2	1	40	1,3	3	14	1,3
3	1,5	49	1,6	6	40	1,6
4	2	52	1,5	9	79	1,5
5	2,5	63	1,4	12	125	1,4
6	3	77	1,3	15	188	1,3
7	3,5	89	1,6	18	257	1,6

Послідовність виконання:

1. В першому наближенні прийняти лінійний вид регресії $y = b_0 + b_1 \cdot x$ (прийнявши в (2.34) m=1).

2. За допомогою МНК визначити параметри лінійної регресії.

3. Провести перевірку адекватності прийнятого виду регресійної моделі за критерієм Фішера (2.38).

4. При невиконанні умови перевірки на адекватність виконати наступне наближення, прийнявши в (2.34) m= m +1.

5. Для використаних моделей та експериментальних даних побудувати графіки.

Приклад розв'язання Завдання 9. Результати експериментальних досліджень представлено в табл. 2.13 (стовбці 2, 3).

В першому наближенні в (2.34) прийнято m=1 – припущення лінійної регресії виду $\tilde{y} = b_0 + b_1 \cdot x$. За МНК підібрано коефіцієнти регресійної моделі $\tilde{y}_1 = -21{,}57 + 15{,}32 \cdot x$. Відхилення між вибірковими даними та розрахованими за регресійною моделлю значень наведено в табл. 2.13. За (2.37) для лінійної моделі $S^2_{ад}(y) = \frac{354{,}53}{7-1} = 59{,}09$.

Значення критерію Фішера F=59,09/ 0,86=68,59.

Критичне значення критерію Фішера F_{1-p}(6; 14; 5%)=2,9.

Перевірка умови (2.38): 68,59 > 2,9 – умова не виконується, отже, гіпотеза про адекватність лінійної регресійної моделі на рівні значимості 0,05 відхиляється.

Таблиця 2. 13

№ з/п	x	y	$S_i^2(y)$	$\tilde{y}_1 = 15{,}32 \cdot x - 21{,}57$		
				$\tilde{y}_1$	$y - \tilde{y}_1$	$(y - \tilde{y}_1)^2$
1	2	3	4	5	6	7
1	1	5	0,92	-6,25	11,25	126,56
2	2	8	0,85	9,07	-1,07	1,14
3	3	17	0,78	24,39	-7,39	54,61
4	4	31	0,81	39,71	-8,71	75,86
5	5	52	0,91	55,03	-3,03	9,18
6	6	70	0,86	70,35	-0,35	0,12
7	7	95	0,9	85,67	9,33	87,05
	Сер.	39,71	0,86			
	$\sum$					354,53

В другому наближенні при припущенні поліноміальної регресії $\tilde{y} = b_0 + b_1 \cdot x + b_2 \cdot x^2$ (в (2.34) прийнято m=2) за МНК підібрано регресійну модель $\tilde{y}_2 = 2{,}57 - 0{,}77 \cdot x + 2{,}01 \cdot x^2$. Розрахунки для квадратичної моделі наведено в табл. 2.14.

За (2.37) для даної моделі $S^2_{ад}(y) = \frac{14{,}53}{7-1} = 2{,}42$.

Значення критерію Фішера F=2,42/ 0,86=2,81.

Перевірка умови (2.38): 2,81 < 2,9 – умова виконується, отже, гіпотеза про адекватність поліноміальної регресійної моделі другого ступеня на рівні значимості 0,05 приймається.

Графіки регресійних моделей наведено на рис. 2.4.

<u>Відповідь:</u> експериментальні дані адекватно описуються квадратичною моделлю $\tilde{y}_2 = 2{,}57 - 0{,}77 \cdot x + 2{,}01 \cdot x^2$ на рівні значимості 0,05.

Таблиця 2. 14

№ з/п	x	y	$S_i^2(y)$	$\tilde{y}_2 = 2{,}01 \cdot x^2 - 0{,}77 \cdot x + 2{,}57$		
				$\tilde{y}_2$	$y - \tilde{y}_2$	$(y - \tilde{y}_2)^2$
1	2	3	4	5	6	7
1	1	5	0,92	3,81	1,19	1,42
2	2	8	0,85	9,07	-1,07	1,14
3	3	17	0,78	18,35	-1,35	1,82
4	4	31	0,81	31,65	-0,65	0,42
5	5	52	0,91	48,97	3,03	9,18
6	6	70	0,86	70,31	-0,31	0,10
7	7	95	0,9	95,67	-0,67	0,45
	Сер.	39,71	0,86			
	$\sum$					14,53

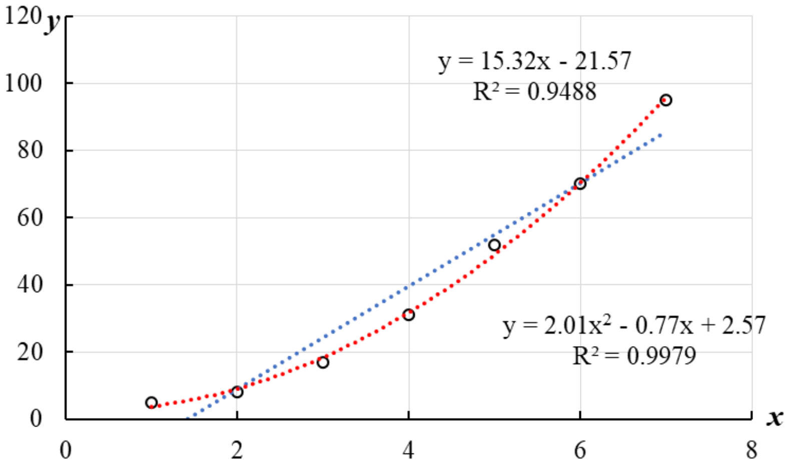


Рис. 2. 4. Графіки регресійних моделей

3. Рекомендована література

Основна

1. Бахрушин В. Є. Математичне моделювання : навч. посіб. / Запоріз. ін-т держ. та муніцип. упр. Запоріжжя : Гуманіт. Ун-т «ЗІДМУ», 2004. 140 с.
2. Бахрушин В. Є. Методи аналізу даних : навчальний посібник. Запоріжжя, 2011.
3. Гідротехнічні споруди : навч. посіб. / М. М. Хлапук, Л. А. Шинкарук, А. В. Дем'янюк, О. А. Дмитрієва. Рівне : НУВГП, 2013. 241 с.
4. Лаврик В. І. Методи математичного моделювання в екології : навч. посібник. К. : Вид. Дім «КМ Академія», 2002. 203 с.
5. Лопотко О. В. Математичні методи в розрахунках на ЕОМ : навч. посіб. /2-ге вид., стереотип. Львів : Магнолія 2006, 2007.
6. Чуйко Г. П., Дворник О. В., Яремчук О. М. Математичне моделювання систем і процесів : навчальний посібник. Миколаїв, 2015.

Допоміжна

1. Kuhn M., Johnson K. (2013) Applied Predictive Modeling, Spr. Sc.+Bus. Media, NY.
2. Berthold M., Borgelt Ch., Höppner F., Klawonn F. (2010) Guide to intelligent data analysis: how to intelligently make sense of real data, Springer-Verlag, London, 2010.

Додаток 1

Квантилі розподілу максимального відносного відхилення v_{1-p}

n	Рівень значимості *1-p*		n	Рівень значимості *1-p*	
	0,10	0,05		0,10	0,05
3	1,406	1,412	15	2,326	2,493
4	1,645	1,689	16	2,354	2,523
5	1,731	1,869	17	2,380	2,551
6	1,894	1,996	18	2,404	2,557
7	1,974	2,093	19	2,426	2,600
8	2,041	2,172	20	2,447	2,623
9	2,097	2,237	21	2,467	2,644
10	2,146	2,294	22	2,486	2,664
11	2,190	2,383	23	2,504	2,683
12	2,229	2,387	24	2,520	2,701
13	2,264	2,425	25	2,537	2,717
14	2,297	2,461			

Додаток 2

Квантилі розподілу Стьюдента

n	Значення *p*				
	0,6	0,8	0,95	0,99	0,999
2	1,376	3,078	12,706	63,657	636,61
3	1,061	1,886	4,303	9,925	31,598
4	0,978	1,638	3,182	5,841	12,941
5	0,941	1,533	2,776	4,604	8,610
6	0,920	1,476	2,571	4,032	6,859
7	0,906	1,440	2,447	3,707	5,959
8	0,896	1,415	2,365	3,499	5,405
9	0,889	1,397	2,306	3,355	5,041
10	0,883	1,383	2,262	3,250	4,781
15	0,868	1,345	2,145	2,977	4,140
20	0,861	1,328	2,093	2,861	3,883
30	0,854	1,311	2,045	2,756	3,659
40	0,851	1,303	2,021	2,704	3,551

Додаток 3

Критичні значення критерію Колмогорова-Смирнова $\frac{K_{n,\alpha}}{\sqrt{n}}$

$n \setminus \alpha$	0,001	0,01	0,02	0,05	0,1	0,15	0,2
1		0,99500	0,99000	0,97500	0,95000	0,92500	0,90000
2	0,97764	0,92930	0,90000	0,84189	0,77639	0,72614	0,68377
3	0,92063	0,82900	0,78456	0,70760	0,63604	0,59582	0,56481
4	0,85046	0,73421	0,68887	0,62394	0,56522	0,52476	0,49265
5	0,78137	0,66855	0,62718	0,56327	0,50945	0,47439	0,44697
6	0,72479	0,61660	0,57741	0,51926	0,46799	0,43526	0,41035
7	0,67930	0,57580	0,53844	0,48343	0,43607	0,40497	0,38145
8	0,64098	0,54180	0,50654	0,45427	0,40962	0,38062	0,35828
9	0,60846	0,51330	0,47960	0,43001	0,38746	0,36006	0,33907
10	0,58042	0,48895	0,45662	0,40925	0,36866	0,34250	0,32257
11	0,55588	0,46770	0,43670	0,39122	0,35242	0,32734	0,30826
12	0,53422	0,44905	0,41918	0,37543	0,33815	0,31408	0,29573
13	0,51490	0,43246	0,40362	0,36143	0,32548	0,30233	0,28466
14	0,49753	0,41760	0,38970	0,34890	0,31417	0,29181	0,27477
15	0,48182	0,40420	0,37713	0,33760	0,30397	0,28233	0,26585
16	0,46750	0,39200	0,36571	0,32733	0,29471	0,27372	0,25774
17	0,45440	0,38085	0,35528	0,31796	0,28627	0,26587	0,25035
18	0,44234	0,37063	0,34569	0,30936	0,27851	0,25867	0,24356
19	0,43119	0,36116	0,33685	0,30142	0,27135	0,25202	0,23731
20	0,42085	0,35240	0,32866	0,29407	0,26473	0,24587	0,23152
25	0,37843	0,31656	0,30349	0,26404	0,23767	0,22074	0,20786
30	0,34672	0,28988	0,27704	0,24170	0,21756	0,20207	0,19029
35	0,32187	0,26898	0,25649	0,22424	0,20184	0,18748	0,17655
40	0,30169	0,25188	0,23993	0,21017	0,18939	0,17610	0,16601
45	0,28482	0,23780	0,22621	0,19842	0,17881	0,16626	0,15673
50	0,27051	0,22585	0,21460	0,18845	0,16982	0,15790	0,14886
> 50	1,94947	1,62762	1,51743	1,35810	1,22385	1,13795	1,07275

Додаток 4
Квантилі розподілу Кохрана

Рівень значимості 0,05														
k	Кількість ступенів вільності *f*													
	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	9985	9750	9392	9057	8772	8534	8332	8159	8010	7880	7341	6602	5813	5000
3	9669	8709	7977	7457	7071	6771	6530	6333	6167	6025	5466	4748	4031	3333
4	9065	7679	6841	6287	5895	5598	5365	5175	5017	4884	4366	3720	3093	2500
5	8412	6838	5981	5441	5065	4783	4564	4387	4241	4118	3645	3066	2513	2000
6	7808	6161	5321	4803	4447	4184	3980	3817	3682	3568	3135	2612	2119	1667
7	7271	5612	4800	4307	3974	3726	3535	3384	3259	3154	2756	2278	1833	1429
8	6798	5157	4377	3910	3595	3362	3185	3043	2926	2829	2462	2022	1616	1250
9	6385	4775	4027	3584	3286	3067	2901	2768	2659	2568	2226	1820	1446	1111
10	6020	4450	3733	3311	3029	2823	2666	2541	2439	2353	2032	1655	1308	1000
20	3894	2705	2205	1921	1735	1602	1501	1422	1357	1303	1108	0879	0675	0500
60	1737	1131	0895	0765	0682	0623	0583	0552	0520	0497	0411	0316	0234	0167

Додаток 5

Критичні значення коефіцієнтів кореляції Пірсона

Кількість ступенів вільності	Рівень значимості для одностороннього критерію	
	0,05	0,01
1	0,9877	0,9995
2	0,9000	0,980
3	0,8054	0,934
4	0,7293	0,882
5	0,6694	0,833
6	0,6215	0,789
7	0,5822	0,750
8	0,5494	0,715
9	0,5214	0,685
10	0,4973	0,658
20	0,3598	0,492
30	0,2960	0,409
60	0,2108	0,295
120	0,1500	0,2100
∞	0,0730	0,1030

Додаток 6

Квантилі *F* - розподілу Фішера рівня значимості 0,05

f_2	f_1													
	1	2	3	4	5	6	7	8	9	10	20	60	120	∞
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	249,05	252,20	253,25	254,32
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396	19,454	19,479	19,487	19,496
3	10,128	9,5521	9,2766	9,1172	9,0135	8,9406	8,8868	8,8452	8,8123	8,7855	8,6385	8,5720	8,5494	8,5265
4	7,7086	6,9443	6,5914	6,3883	6,2560	6,1631	6,0942	6,0410	5,9988	5,9644	5,7744	5,6878	5,6581	5,6281
5	6,6079	5,7861	5,4095	5,1922	5,0503	4,9503	4,8759	4,8183	4,7725	4,7351	4,5272	4,4314	4,3984	4,3650
6	5,9874	5,1433	4,7571	4,5337	4,3874	4,2839	4,2066	4,1468	4,0990	4,0600	3,8415	3,7398	3,7047	3,6688
7	5,5914	4,7374	4,3468	4,1203	3,9715	3,8660	3,7870	3,7257	3,6767	3,6365	3,4105	3,3043	3,2674	3,2298
8	5,3177	4,4590	4,0662	3,8378	3,6875	3,5806	3,5005	3,4381	3,3881	3,3472	3,1152	3,0053	2,9669	2,9276
9	5,1174	4,2565	3,8626	3,6331	3,4817	3,3738	3,2927	3,2296	3,1789	3,1373	2,9005	2,7872	2,7475	2,7067
10	4,9646	4,1028	3,7083	3,4780	3,3258	3,2172	3,1355	3,0717	3,0204	2,9782	2,7372	2,6211	2,5801	2,5379
20	4,3513	3,4928	3,0984	2,8661	2,7109	2,5990	2,5140	2,4471	2,3928	2,3479	2,0825	1,9464	1,8963	1,8432