

Лекція 6. Обробка результатів наукових досліджень

План лекції:

1. Обробка результатів експериментальних досліджень. Основи теорії випадкових помилок та методів оцінки випадкових похибок у вимірюваннях.
- 6.2. Методи графічної обробки результатів експерименту.
- 6.3. Аналітична обробка результатів експерименту.
- 6.4. Елементи теорії планування експерименту.

1. Обробка результатів експериментальних досліджень. Основи теорії випадкових помилок та методів оцінки випадкових похибок у вимірюваннях

Аналіз випадкових похибок базується на теорії випадкових помилок, яка дає можливість з визначеною гарантією обчислити дійсне значення вимірюваної величини та оцінити можливі помилки.

Основу теорії випадкових помилок складають такі **припущення**:

- при великій кількості вимірів випадкові похибки однакової величини, але різного знаку зустрічаються однаково часто;
- більші похибки зустрічаються рідше, ніж малі (випадковість появи похибки зменшується зі зростанням її величини);
- при нескінченно великому числі вимірів істинне значення вимірюваної величини дорівнює середньоарифметичному значенню всіх результатів вимірів, а поява того чи іншого результату вимірів як випадкової події описується нормальним законом розподілу.

Вирізняють **генеральну та вибіркову сукупність вимірів**.

Під **генеральною сукупністю** розуміють всю множину можливих значень вимірів x_i або можливих значень похибок Δx_i .

Для **вибіркової сукупності** число вимірів n обмежене і у кожному конкретному випадку строго визначається.

Звичайно вважають, що якщо $n > 30$, то середнє значення даної сукупності вимірів $\bar{x}$ достатньо наближається до його дійсного значення.

Теорія випадкових похибок дозволяє оцінити точність та надійність вимірів при даній кількості вимірів або визначити мінімальну кількість вимірів, що гарантує задану точність та надійність вимірів. Разом з цим виникає необхідність виключити грубі похибки ряду, визначити достовірність одержаних даних тощо.

Для великої вибірки та нормального закону розподілу загальною оціночною характеристикою вимірів є дисперсія D та коефіцієнт варіації κ_σ :

$$D = \sigma^2 = \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] / n; \quad \hat{\sigma}_a = \sigma / \bar{x}. \quad (6.1)$$

Дисперсія характеризує однорідність вимірів. Чим вище D , тим більше розкид вимірів. Чим вище κ_σ , тим більше мінливість вимірів відносно середніх значень, κ_σ оцінює також розкид при оцінці кількох вибірок.

Довірчим інтервалом значень x_i є такий інтервал, в який потрапляє дійсне значення x_d вимірюваної величини із заданою імовірністю.

Довірчою імовірністю (вірогідністю) (P_d) вимірів називається імовірність того, що істинне значення вимірюваної величини потрапить до даного довірчого інтервалу, тобто в зону $a \leq x_d \leq b$. Ця величина визначається у частках одиниці або у відсотках.

Довірчий інтервал характеризує точність вимірів даної вибірки, а довірча імовірність – достовірність виміру.

Довірча імовірність визначається за допомогою інтегральної функції Лапласа, а довірчий інтервал визначається при $n > 30$ за допомогою аргументу функції Лапласа, а при $n < 30$ - за допомогою аргументу функції Стюдента.

Для проведення дослідів із заданою точністю та достовірністю необхідно знати ту кількість вимірів, при якій експериментатор буде впевнений у позитивному результаті. У зв'язку з цим одним із найперших завдань при статистичних методах оцінки є встановлення мінімального, але достатнього числа вимірів для даних умов. Завдання зводиться до встановлення мінімального обсягу вибірки (числа вимірів або спостережень) N_{min} при заданих значеннях довірчого інтервалу та довірчій імовірності.

Для визначення N_{min} може бути використана така послідовність розрахунків:

1) проводиться попередній експеримент з кількістю вимірів n , які становлять залежно від трудомісткості дослідів від 20 до 50;

2) розраховується середньоквадратичне відхилення σ ;

3) відповідно до поставлених завдань експерименту визначається потрібна точність вимірів Δ за формулою

$$\Delta = \sigma_0 / \bar{x}, \quad (6.2)$$

де σ_0 - середньоарифметичне значення середньоквадратичного відхилення σ , яке дорівнює $\sigma_0 = \sigma / \sqrt{n}$.

4) встановлюється нормоване відхилення t , значення якого звичайно задається (залежить також від точності методу);

5) визначають N_{min} за такою формулою:

$$N_{min} = \sigma^2 t^2 / \sigma_0^2 = k_a^2 t^2 / \Delta^2, \quad (6.3)$$

де k_a – коефіцієнт варіації (мінливості), %;

Δ – точність вимірів, %.

При подальшому проведенні експерименту число вимірів не повинне бути меншим за N_{min} .

У процесі обробки експериментальних даних слід виключати грубі помилки ряду. Однак перш ніж виключити той чи інший вимір, необхідно упевнитись, що це дійсно помилка, а не відхилення внаслідок статистичного розкиду. Найпростішим способом виключення із ряду виміру, що різко відрізняється від інших, є **правило трьох сигм**: розкид випадкових величин від середнього значення не повинен перевищувати

$$x_{max,min} = \bar{x} \pm 3\sigma. \quad (6.4)$$

Більш достовірним є метод, що базується на використанні довірчого інтервалу. Другим методом встановлення грубих помилок є метод, що

базується на використанні критерію В.І. Романовського. Ці методи використовують за наявності малої вибірки.

У випадку більш глибокого аналізу експериментальних даних рекомендується така послідовність:

1) після одержання експериментальних даних у вигляді статистичного ряду його аналізують і виключають систематичні помилки;

2) аналізують ряд з метою виявлення грубих помилок та похибок: встановлюють підозрілі значення $x_{\max}$ або $x_{\min}$; визначають величину середньоквадратичного відхилення σ ; розраховують критерії виключення із статистичного ряду значень $x_{\max}$ та $x_{\min}$ (за допомогою одного з двох згаданих вище методів); виключають за необхідності із статистичного ряду $x_{\max}$ та $x_{\min}$ і одержують новий ряд із нових членів;

3) розраховують середньоарифметичне $\bar{x}$, похибки окремих вимірів $(\bar{x} - x_i)$ та середньоквадратичне очищеного ряду σ ;

4) знаходять середньоквадратичне σ_0 серії вимірів, коефіцієнт варіації k_v ;

5) при великій вибірці задаються довірчою імовірністю $p_A = \varphi(t)$ або рівнянням значущості $(1 - P_D)$ та за допомогою таблиць значень інтегральної функції Лапласа визначають t ;

6) визначають довірчий інтервал μ_{c0} ;

7) встановлюють дійсне значення величини, що досліджується за формулою

$$x_A = \bar{x} \pm \mu_{c0}; \quad (6.5)$$

8) оцінюють відносну похибку результатів серії вимірів при заданій довірчій імовірності P_D .

6.2. Методи графічної обробки результатів експерименту

При обробці результатів вимірів та спостережень широко використовують **методи графічного зображення**. Графічне зображення надає найбільш наочне уявлення про результати експерименту, дозволяє краще пізнати фізичну сутність досліджуваного процесу, виявити загальний характер функціональної залежності змінних величин, що вивчаються, встановити наявність максимуму або мінімуму функції.

Для графічного зображення результатів досліджень, як правило, використовують систему прямокутних координат. Якщо аналізується графічним методом функція $y=f(x)$, то наносять в системі прямокутних координат значення $x_1 y_1, x_2 y_2, \dots, x_n y_n$. Перш ніж будувати графік, необхідно знати хід (перебіг) досліджуваного явища. Як правило, експериментатор має орієнтовне уявлення про якісні закономірності та форму графіка з теоретичних досліджень.

Точки на графіку необхідно з'єднувати плавною лінією таким чином, щоб вона проходила найближче до всіх експериментальних точок. Якщо з'єднати точки прямими відрізками, то одержимо ламану криву. Вона характеризує зміну функції за даними експерименту. Звичайно функції мають плавний характер. Тому при графічному зображенні результатів дослідження слід

проводити між точками плавні криві. Різке викривлення графіка пояснюється похибками вимірів. Якби експеримент повторили з використанням засобів вимірювання більш високої точності, то одержали б менші похибки, а ламана крива більше відповідала б плавній кривій.

Однак можуть бути і винятки, тому що іноді досліджуються явища, для яких у визначених інтервалах спостерігається швидка стрибкоподібна зміна однієї з координат. У таких випадках необхідно особливо ретельно з'єднувати точки кривої.

Іноколи при побудові графіка спостерігається різке віддалення однієї або двох точок від кривої. У цьому випадку спочатку потрібно проаналізувати фізичну сутність явища, і якщо немає підстав до стрибку функції, то таке різке відхилення можна пояснити грубою похибкою [1, 8].

Часто при графічному зображенні результатів експерименту доводиться мати справу з трьома змінними $b=f(x, y, z)$. У цьому випадку використовують метод поділу змінних. Одній з величин z у межах інтервалу вимірів (z_1 - z_n) надають кілька послідовних значень. Для двох інших змінних x та y будують графіки функцій $y=f_i(x)$ при $z_i = \text{const}$. У результаті на одному й тому самому графіку одержують сімейство кривих $y=f_i(x)$ для різних значень z .

При графічному зображенні результатів експерименту великого значення набуває вибір масштабів та координатної сітки.

Координатні сітки бувають рівномірні та нерівномірні. У рівномірних координатних сітках ординати та абсциси мають рівномірну шкалу.

Нерівномірна координатна сітка використовується для більшої наочності у випадках, коли функція має різко змінюваний характер. З нерівномірних координатних сіток найбільшого поширення набули **напівлогарифмічні, логарифмічні та імовірнісні**.

Напівлогарифмічна сітка має рівномірну шкалу на ординаті та логарифмічну шкалу на абсцисі.

Логарифмічна координатна сітка має на двох осях логарифмічні шкали.

Імовірнісні координатні сітки мають на ординаті, як правило, рівномірну шкалу, а на осі абсцис - імовірнісну шкалу.

Доцільність використання нерівномірної функціональної сітки пояснюється, крім вищезазначеного, бажанням представити функцію, що досліджується, у вигляді прямої, що підвищує точність побудови. При цьому здійснюється так зване вирівнювання, тобто криву, що побудована за дослідними даними, представляють лінійною функцією. Нехай, наприклад, для деякої емпіричної кривої підібрана функція типу $y = ax^n$. Процес вирівнювання буде таким. Наведений вираз перетворюється за допомогою логарифмування у вираз $\lg y = n \lg x + \lg a$. Якщо ми позначимо $\lg y = y_1; \lg x = x_1; \lg a = a_1$, то одержимо лінійну формулу $y_1 = nx_1 + a_1$, графіком якої буде пряма лінія.

Масштаб за координатними осями, як правило, використовують різний. Від його вибору залежить форма графіка – він може бути пласким (вузьким) або витягнутим (широким) вдовж осі. Вузькі графіки дають більшу похибку на осі ординат, широкі на осі абсцис. Правильно підібраний масштаб дозволяє підвищити точність відрахування. Розрахункові графіки, що мають екстремум функції або який-небудь складний вигляд, особливо ретельно потрібно

креслити у зонах вигину. На таких ділянках кількість точок для накреслення графіка повинно бути істотно більшою, ніж на плавних ділянках.

У деяких випадках будують **номограми**, які істотно полегшують використання для систематичних розрахунків складних теоретичних та емпіричних формул у відповідних межах зміни величин.

Номограма (від грец. **nomos** – закон та **gramma** – риска, буква, писемний знак, зображення) – креслення, яке є зображенням функціональних залежностей, які використовуються для одержання (без розрахунків) приблизних розв'язань рівнянь.

6.3. Аналітична обробка результатів експерименту

У процесі експериментальних вимірів звичайно одержують статистичний ряд вимірів двох величин, які об'єднуються функцією $y=f(x)$. Кожному значенню функції $y_1, \dots, y_n$ відповідає відповідне значення аргументу $x_1, x_2, \dots, x_n$.

На основі експериментальних даних можна підібрати алгебраїчні вирази, які називають емпіричними формулами. Такі формули підбирають лише у межах вимірюваних значень аргументу $x_1 - x_n$. Емпіричні формули мають тим більшу цінність, чим більше вони відповідають результатам експерименту. Досвід показує, що емпіричні формули є незамінними для аналізу вимірюваних величин. До емпіричних формул висувають дві основні вимоги – по можливості вони повинні бути найбільш простими та точно відповідати експериментальним даним у межах зміни аргументу. Таким чином емпіричні формули є приблизним виразом аналітичних. Заміну точних аналітичних виразів приблизними, більш простими називають апроксимацією, а функції апроксимуючими.

Процес підбору емпіричних формул складається з двох етапів. На першому етапі дані вимірів наносять на сітку прямокутних координат, поєднують експериментальні точки плавною кривою і вибирають орієнтовно вид формули. На другому етапі обчислюють параметри формул, які найкраще відповідали б прийнятій формулі. Підбір емпіричних формул необхідно починати з найбільш простих виразів.

Криві, що побудовані за експериментальними точками, вирівнюють за допомогою статистичних методів. Наприклад, методом вирівнювання, який полягає в тому, що криву, яку побудовано за експериментальними точками, представляють лінійною функцією. Для знаходження параметрів заданих рівнянь часто використовують метод середніх та найменших квадратів.

Для дослідження закономірностей між явищами (процесами), які залежать від багатьох факторів, використовують кореляційний аналіз.

У процесі проведення експерименту виникає потреба перевірити відповідність експериментальних даних теоретичним передумовам, тобто перевірити гіпотезу дослідження. Перевірка експериментальних даних на адекватність необхідна також у всіх випадках на стадії аналізу теоретико-експериментальних досліджень. У практиці адекватності використовують різні критерії: Фішера, Пірсона, Романовського.

6.4. Елементи теорії планування експерименту

Теорія математичного експерименту включає ряд концепцій, які забезпечують успішну реалізацію завдань дослідження. До них відносять *концепції рандомізації, послідовного експерименту, математичного моделювання, оптимального використання факторного простору* і деякі інші.

Принцип рандомізації полягає в тому, що до плану експерименту вводять елемент імовірності. Для цього план експерименту складають так, щоб ті систематичні фактори, які складно піддаються контролю, враховувалися статистично і потім виключалися в дослідженнях як систематичні похибки.

При *послідовному проведенні експерименту* виконується не одночасно, а поетапно, для того щоб результати кожного етапу аналізувати та приймати рішення про доцільність проведення подальших досліджень.

У результаті експерименту одержують рівняння регресії, яке часто називають *математичною моделлю процесу*. Для конкретних випадків математична модель створюється на основі цільової направленості процесу та завдань дослідження з урахуванням визначеної точності рішення та достовірності вихідних даних, яка звичайно проводиться за критерієм Фішера. В зв'язку з тим, що ступінь полінома, який адекватно описує процес, передбачити неможливо, спочатку намагаються описати явище лінійною моделлю, а потім, якщо вона неадекватна, підвищують ступінь полінома тобто проводять експеримент поетапно.

Важливе місце в теорії планування експерименту займають питання *оптимізації* процесів, що досліджуються якостей багатокомпонентних систем або інших об'єктів.

Як правило, не можна знайти таке поєднання значень факторів впливу, при якому одночасно досягається екстремум всіх функцій відгуку.

Тому у більшості випадків за критерій оптимальності вибирають лише одну зі змінних стану – функцію відгуку, що характеризує процес, а інші беруть прийнятними для даного випадку.