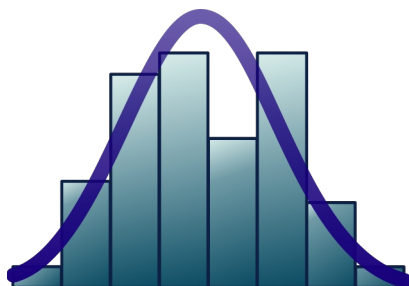


Міністерство освіти і науки України
Чернівецький національний університет
імені Юрія Федьковича

Спеціальні глави вищої математики

Методичні вказівки до курсової роботи



Чернівці
ЧНУ ім. Юрія Федьковича
2022

УДК 519.2(072)
С718

*Рекомендовано Вченою радою
навчально-наукового Інституту фізико-технічних
та комп'ютерних наук
Чернівецького національного університету
імені Юрія Федьковича
(Протокол № 1 від 26.01.2022)*

Укладачі: Івашко Віктор Вікторович, канд. фіз.-мат. наук, асистент.
Фельде Христина Вікторівна, канд. фіз.-мат. наук, доцент.

С718 Івашко В.В., Фельде Х.В. Спеціальні глави вищої математики :
методичні вказівки до курсової роботи. Чернівці : Чернівецький
національний університет імені Юрія Федьковича, 2022. 34 с.

У методичній розробці наведені вказівки до курсової роботи з
дисципліни «Спеціальні глави вищої математики».

Для студентів технічних факультетів та інститутів за
спеціальністю «Метрологія та інформаційно-вимірювальна техніка».

УДК 519.2(072)
© Чернівецький національний університет
імені Юрія Федьковича, 2022

ЗМІСТ

| | |
|---|----|
| ВСТУП | 4 |
| Вимоги до курсової роботи..... | 5 |
| Критерії оцінювання якості курсової роботи..... | 6 |
| Завдання до курсової роботи..... | 8 |
| Теоретичні відомості..... | 10 |
| Розрахунок числових характеристик вибірки..... | 13 |
| Графічне представлення..... | 16 |
| Побудова емпіричної функції розподілу..... | 19 |
| Побудова графіка рівняння лінійної регресії..... | 21 |
| Побудова графіка рівняння криволінійної регресії..... | 23 |
| Розрахунок довірчих інтервалів..... | 25 |
| Виконання прогнозу..... | 27 |
| Висновки до курсової роботи..... | 28 |
| СПИСОК ЛІТЕРАТУРИ..... | 29 |
| ДОДАТОК А..... | 30 |
| ДОДАТОК Б..... | 31 |
| ДОДАТОК В..... | 32 |

ВСТУП

Дані методичні вказівки мають на меті надати студентам основні рекомендації та показати алгоритм дій щодо виконання курсової роботи з навчальної дисципліни «Спеціальні глави вищої математики».

Основний акцент робиться на розвитку у студентів навичок самостійно приймати рішення при здійсненні ними статистичного аналізу випадкових величин, тобто утворення вибірок цих величин, розрахунок числових характеристик, виконання оцінки, графічного представлення та прогнозу.

Методичні вказівки містять перелік завдань і приклади для їх виконання, які можна використовувати при здійсненні аналізу функціональної залежності від часу будь-якої дискретної випадкової величини та виконанні точкового прогнозу.

Вимоги до курсової роботи

Курсова робота може бути рукописною (написаною розбірливим почерком) або надрукованою на принтері з однієї сторони аркуша.

Роботу оформлюють на аркушах білого паперу формату А4 (21 × 29,7 см). Розміри полів: зверху – 2 см, знизу – 2,5 см, зліва – 3 см, справа – 1,5 см. Шрифт – Times New Roman або Arial, розмір – 12-14 пунктів. Міжрядковий інтервал – 1,5.

Нумерація сторінок здійснюється на колонтитулах, номер сторінки проставляється у нижньому (або верхньому) правому куті або по центру сторінки арабськими цифрами. Крапка після цифри не ставиться. Титульний аркуш не нумерується. Кожна частина роботи (розділ) починається з нової сторінки.

Рекомендований обсяг курсової роботи – 25-30 сторінок. Цей обсяг може відхилитися в межах $\pm 10\%$.

Структура курсової роботи будується відповідно до мети, завдання та вимог. Рекомендована структура курсової роботи:

- титульний аркуш;
- завдання до роботи (за загальною формою);
- реферат;
- зміст;
- список умовних позначень, термінів, скорочень (за необхідності);
- вступ;
- основна частина;
- висновки;
- список літератури;
- додатки (за необхідності).

Критерії оцінювання якості курсової роботи

Відповідно до завдання, студент виконує всі розділи курсової роботи згідно змісту. Захищається курсова робота при комісії (2-3 викладача).

Остаточна оцінка за курсову роботу виставляється за результатами її захисту в присутності комісії. Захист роботи оцінюється за шкалою ECTS та національною шкалою.

При виставленні остаточної оцінки за курсову роботу члени комісії зі захисту курсової роботи обов'язково враховують такі моменти:

- Відповідність роботи вимогам її написання;
- Оцінку керівника курсової роботи;
- Захист роботи її автором перед членами комісії.

Шкала оцінювання курсової роботи

| Оцінка за шкалою ECTS | Критерії оцінок | Оцінка в балах | Оцінка за національною шкалою |
|-----------------------|---|----------------|-------------------------------|
| A | Студент показав вільне і глибоке володіння змістом курсової роботи. Точно і повно відповів на всі запитання членів комісії. Вільно володіє термінологією. | 90-100 | Відмінно |
| B | Студент показав глибоке володіння змістом курсової роботи. При відповіді на запитання студентом були допущені незначні неточності, які він зумів повністю виправити після того, як на них було звернуто увагу з боку членів комісії. В основному володіє термінологією. | 80-89 | Добре |

| | | | |
|----|--|-------|--------------|
| C | Студент показав, що він в основному володіє змістом курсової роботи. При відповіді на запитання студентом були допущені незначні неточності, які він не зумів повністю виправити після того, як на них було звернуто увагу з боку членів комісії. В основному володіє термінологією. | 70-79 | Добре |
| D | Студент показав, що він в основному володіє змістом курсової роботи. Відповіді на запитання членів комісії були не зовсім чітко сформульовані. Деякі терміни студент використовує не за їх призначенням. | 60-69 | Задовільно |
| E | Студент показав, що він в основному володіє змістом курсової роботи. Відповіді на запитання членів комісії були нечітко сформульовані. Деякі терміни студент використовує не за їх призначенням. | 50-59 | Задовільно |
| FX | Студент показав, що він не володіє частиною змісту курсової роботи. Відповіді на запитання членів комісії є поверхневими. Ряд висновків є неправильно обґрунтовуваними. Знання термінів незадовільне. | 35-49 | Незадовільно |
| F | Студент показав, що він не володіє частиною змісту курсової роботи. Відповіді на запитання членів комісії є нечіткими. Ряд висновків є неправильними. Знання термінів незадовільне. | 1-34 | Незадовільно |

Завдання до курсової роботи

Мета курсової роботи полягає в проведенні аналізу функціональної залежності від часу дискретної випадкової величини та виконанні точкового прогнозу, що здійснюється у рамках теорії ймовірностей та математичної статистики.

Завдання до курсової роботи:

- вибір залежної від часу дискретної випадкової величини (метрологічна або тому подібна);
- утворення вибірки, яка повинна містити як мінімум 100 пар чисел (X, Y) ;
- розрахунок параметрів вибірки (вибіркове середнє, дисперсія, виправлена дисперсія, середньоквадратичне відхилення, виправлене середньоквадратичне відхилення та вибірковий коефіцієнт кореляції);
- графічне представлення вибірки (діаграма розсіювання, полігон та гістограма частот);
- побудова емпіричної функцію розподілу за Y ;
- побудова графіку рівняння лінійної регресії Y на X ;
- побудова графіку рівняння криволінійної регресії Y на X ;
- побудова довірчих інтервалів для математичного сподівання Y , виконання прогнозу та висновки.

Як *приклад* проаналізуємо зміну кількості витрат електроенергії (кВт·год) на базі електролічильника «NIK 2102-02 M2B (5-60A)» за період 2013-2021 років (з 01.09.2013 по 31.12.2021 включно) та виконаємо точковий прогноз.

Утворимо вибірку (об'ємом $n = 100$) залежності величини Y від X , де X – це місяці та Y – кількість спожитої електроенергії (кВт·год) за місяць (див. табл. 1).

Таблица 1

| X | Y | X | Y | X | Y | X | Y |
|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 302 | 26 | 155 | 51 | 465 | 76 | 422 |
| 2 | 452 | 27 | 290 | 52 | 396 | 77 | 519 |
| 3 | 219 | 28 | 276 | 53 | 183 | 78 | 367 |
| 4 | 910 | 29 | 12 | 54 | 334 | 79 | 407 |
| 5 | 723 | 30 | 315 | 55 | 337 | 80 | 302 |
| 6 | 461 | 31 | 519 | 56 | 250 | 81 | 118 |
| 7 | 686 | 32 | 377 | 57 | 131 | 82 | 127 |
| 8 | 492 | 33 | 402 | 58 | 131 | 83 | 164 |
| 9 | 333 | 34 | 559 | 59 | 220 | 84 | 106 |
| 10 | 150 | 35 | 559 | 60 | 151 | 85 | 167 |
| 11 | 388 | 36 | 432 | 61 | 143 | 86 | 148 |
| 12 | 125 | 37 | 398 | 62 | 158 | 87 | 376 |
| 13 | 150 | 38 | 175 | 63 | 326 | 88 | 549 |
| 14 | 617 | 39 | 306 | 64 | 403 | 89 | 611 |
| 15 | 536 | 40 | 419 | 65 | 338 | 90 | 542 |
| 16 | 391 | 41 | 343 | 66 | 364 | 91 | 381 |
| 17 | 415 | 42 | 62 | 67 | 373 | 92 | 246 |
| 18 | 311 | 43 | 353 | 68 | 142 | 93 | 152 |
| 19 | 266 | 44 | 223 | 69 | 130 | 94 | 126 |
| 20 | 208 | 45 | 209 | 70 | 127 | 95 | 130 |
| 21 | 219 | 46 | 176 | 71 | 123 | 96 | 117 |
| 22 | 286 | 47 | 80 | 72 | 91 | 97 | 145 |
| 23 | 261 | 48 | 196 | 73 | 94 | 98 | 484 |
| 24 | 217 | 49 | 241 | 74 | 139 | 99 | 559 |
| 25 | 217 | 50 | 241 | 75 | 413 | 100 | 663 |

Теоретичні відомості

Нижче наведено основні визначення та терміни з курсу спеціальні глави вищої математики (теорія ймовірностей та математична статистика).

Теорія ймовірностей – розділ математики, який вивчає закономірності випадкових явищ: випадкові події, випадкові величини, їх властивості та операції над ними.

Випадковою величиною називається така величина, яка може набувати різних (випадкових) значень.

Дискретна випадкова величина – це випадкова величина, множина значень якої зліченна (тобто скінченна).

Неперервна випадкова величина – це така випадкова величина, яка може набувати свої можливі значення з деякого скінченного чи нескінченного інтервалу (проміжку).

Математична статистика – наука, яка за допомогою математичних методів встановлює закономірності масових явищ, їх взаємозв'язки, систематизує та обробляє експериментальні дані.

Генеральна сукупність – сукупність об'єктів (одиниць), щодо яких роблять висновки стосовно конкретної проблеми або сукупність об'єктів з яких утворюється вибірка сукупність.

Вибіркова сукупність (вибірка) – частина об'єктів з генеральної сукупності відібраних випадковим чином для аналізу, вивчення та отримання висновків щодо генеральної сукупності.

Об'ємом або *обсягом вибірки* (n) називають кількість об'єктів вибірки.

На практиці використовують різні способи відбору. Проте загалом всі способи можна поділити на два типи:

- відбір, при якому не відбувається поділ генеральної сукупності на частини. Сюди відносять простий випадковий безповторний відбір і простий випадковий повторний відбір;
- відбір, при якому відбувається поділ генеральної сукупності на частини. Сюди відносять типовий, механічний та серійний відбір.

Простим випадковим відбором називають такий відбір при якому об'єкти вилучають по одному з генеральної сукупності.

Типовим називають відбір, при якому об'єкти вилучаються не з усієї генеральної сукупності, а з окремих «типових» її частин. Наприклад, якщо деталі виготовляють на декількох верстатах, то відбір відбувається не зі всієї сукупності деталей, виготовлених всіма верстами, а зі сукупності виготовлених кожним верстатом окремо.

Механічним називають відбір, при якому генеральну сукупність «механічно» ділять на стільки груп, скільки об'єктів повинно увійти у вибірку, а з кожної групи вибирають по одному об'єкту.

Серійним називають відбір, при якому об'єкти з генеральної сукупності вибирають не по одному, а серіями, які підлягають розгляду/дослідженню.

Значення ознаки (вік, професія, вимірювальна величина і т. п.) вибірки, що аналізується називаються *варіантами* (x_i), а послідовність варіантів разом зі їх частотами (n_i), розташованих у порядку зростання, – *варіаційним рядом*.

Для випадкових величин характерними є функціональна (англ. functional) та кореляційна (англ. correlation) залежності.

Під *функціональною* розуміють таку залежність, коли зміна одного фактора зумовлює зміну іншого, при цьому одному значенню незалежного фактора відповідає тільки одне значення залежного фактора.

Кореляція в математичній статистиці – це ймовірнісна (статистична) залежність між величинами, яка не має строго функціонального характеру.

Під *кореляційною* розуміють таку залежність, при якій зміна однієї випадкової величини викликає зміну середнього значення іншої, тобто одному значенню незалежної змінної можуть відповідати кілька значень залежної змінної. Тому кореляційні залежності можуть бути встановлені тільки при обробці великої кількості спостережень.

При роботі з такими залежностями користуються *кореляційним аналізом*, який встановлює кількісну оцінку «щільності» (густини) зв'язку між досліджуваними ознаками (факторами). Щільність зв'язку (наявність кореляції) між двома величинами візуально визначають за допомогою кореляційного поля.

Кореляційне поле – це сукупність точок у прямокутній системі координат, абсциса кожної з яких відповідає значенню факторної ознаки (X), а ордината – значенню результативної ознаки (Y) випадкової величини, що аналізується. Кількість точок на діаграмі відповідає кількості одиниць спостереження.

Коефіцієнт кореляції (r) у математичній статистиці – параметр, який характеризує силу статистичного зв'язку між двома або декількома випадковими величинами. Якщо коефіцієнт кореляції описує зв'язок між двома випадковими величинами його називають *простим* або *коефіцієнтом кореляції Пірсона*.

Розрахунок числових характеристик вибірки

До числових характеристик вибірки належать:

- вибіркове середнє;
- вибіркова дисперсія;
- вибіркове середньоквадратичне відхилення;
- мода вибірки;
- медіана вибірки.

Вибірковим середнім \overline{x}_B називається середнє арифметичне значення ознаки X вибіркової сукупності.

$$\overline{x}_B = \left(\sum_{i=1}^n x_i \right) / n. \quad (3.1)$$

Якщо значення ознаки x_1, x_2, \dots, x_k вибіркової сукупності мають відповідні частоти n_1, n_2, \dots, n_k , причому $n_1 + n_2 + \dots + n_k = n$, то

$$\overline{x}_B = \left(\sum_{i=1}^k n_i x_i \right) / n. \quad (3.2)$$

Вибірковою дисперсією D_B називають середнє арифметичне квадратів відхилення спостережуваних значень від їх середнього \overline{x}_B .

$$D_B = \left(\sum_{i=1}^n (x_i - \overline{x}_B)^2 \right) / n \quad \text{або} \quad D_B = \left(\sum_{i=1}^k n_i (x_i - \overline{x}_B)^2 \right) / n \quad (3.3)$$

Виправлена вибіркова дисперсія обчислюється за формулою:

$$s^2 = \frac{n}{n-1} D_B. \quad (3.4)$$

Вибірковим середньоквадратичним відхиленням σ_B (*стандарт*) називають корінь квадратний із вибіркової дисперсії:

$$\sigma_B = \sqrt{D_B} \quad (3.5)$$

Для оцінки середньоквадратичного відхилення генеральної сукупності використовують «виправлене» середньоквадратичне відхилення, яке дорівнює корінню квадратному із виправленої вибіркової дисперсії:

$$s = \sqrt{\left(\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2 \right) / n - 1}. \quad (3.6)$$

Мода вибірки Mo – те її значення, яке трапляється найчастіше.

Медіана вибірки Me – це число, яке «ділить навпіл» варіаційний ряд. Тобто варіант, який ділить за кількістю варіантів варіаційний ряд на дві рівні між собою частини:

$$Me = \begin{cases} x_m \text{ при } n = 2m - 1; \\ \frac{x_m + x_{m+1}}{2} \text{ при } n = 2m, \end{cases} \quad (3.7)$$

де x_m – середина варіаційного ряду.

Для кількісної характеристики «щільності» або густини кореляційного зв'язку між двома величинами X і Y генеральної сукупності використовують коефіцієнт кореляції.

На практиці для оцінки густини кореляційного зв'язку між величинами X і Y за результатами вибірових спостережень використовують *вибіровий коефіцієнт кореляції*, який знаходиться за допомогою формули:

$$r_B = \frac{\sum n_{xy} xy - n \bar{x}_B \bar{y}_B}{n \sigma_x \sigma_y}, \quad (3.8)$$

де x та y – варіанти ознаки X та Y , n_{xy} – частота появи пар варіантів (x, y) , \bar{x}_B та \bar{y}_B – вибірові середні ознаки X та Y , n – об'єм

вибірки, σ_x та σ_y – вибіркові середньоквадратичні відхилення ознаки X та Y .

Коефіцієнт кореляції є безрозмірною величиною (тому що розмірність чисельника та знаменника є розмірністю добутку X на Y), а його значення не залежить від вибору одиниць виміру обох величин та знаходиться в межах $-1 \leq r_B \leq 1$.

Якщо $r_B = 0$, то говорять, що величини X та Y – незалежні. Якщо $r_B = \pm 1$, то X та Y зв'язані між собою лінійною функціональною залежністю.

Здійснивши розрахунок зазначених вище параметрів вибірки, можна дати відповідну оцінку для генеральної сукупності.

Обчислимо, наприклад, числові характеристики вибірки для ознаки Y та вибірковий коефіцієнт кореляції між X та Y (див. табл. 1).

- Вибіркове середнє: $\bar{y}_B = 305$;
- Вибіркова дисперсія: $D_B = 28911$;
- Виправлена вибіркова дисперсія: $s^2 = 29203$;
- Вибіркове середньоквадратичне відхилення: $\sigma_B = 170$;
- «Виправлене» середньоквадратичне відхилення: $s = 171$;
- Мода вибірки: $Mo = 559$;
- Медіана вибірки: $Me = 302$;
- Вибірковий коефіцієнт кореляції $r_B = -0,2$.

Графічне представлення

Найпоширенішим та найпростішим зі способів графічного представлення вибірки є зображення її значень за допомогою кореляційного поля, тобто побудова діаграми розсіювання (див. рис. 1).

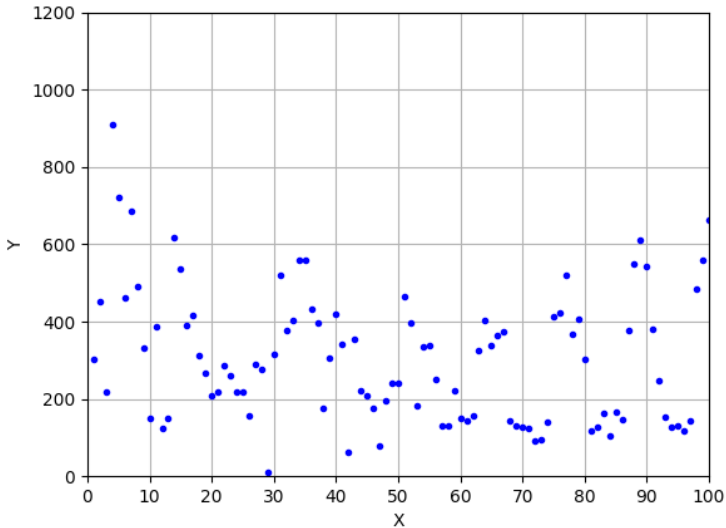


Рис. 1. Кореляційне поле залежності $Y=f(X)$

Для спрощення подальшого статистичного аналізу досліджуваної випадкової величини (Y) утворимо згрупований розподіл частот вибірки.

Алгоритм побудови згрупованого розподілу частот:

- знаходження розмаху варіаційного ряду $R = x_{max} - x_{min}$;
- вибір непарного числа k . При обсягу вибірки $n \geq 100$ доцільно вибирати k з інтервалу $9 \leq k \leq 15$;
- визначення ширини інтервалів варіаційного ряду $h = R/k$;
- встановлення границі інтервалів і визначення частоти варіантів n_i для кожного з них.

Для прикладу, що наводиться, ширина інтервалу $h = 60$.

Згрупований розподіл частот вибірки (Y). Таблиця 2

| № | $[y_i, y_{i+1})$ | y_m | n_i |
|----|------------------|-------|-------|
| 1 | [12, 72) | 42 | 2 |
| 2 | [72, 132) | 102 | 15 |
| 3 | [132, 192) | 162 | 16 |
| 4 | [192, 252) | 222 | 13 |
| 5 | [252, 312) | 282 | 9 |
| 6 | [312, 372) | 342 | 10 |
| 7 | [372, 432) | 402 | 15 |
| 8 | [432, 492) | 462 | 5 |
| 9 | [492, 552) | 522 | 6 |
| 10 | [552, 612) | 582 | 4 |
| 11 | [612, 672) | 642 | 2 |
| 12 | [672, 732) | 702 | 2 |
| 13 | [732, 792) | 762 | 0 |
| 14 | [792, 852) | 822 | 0 |
| 15 | [852, 912) | 882 | 1 |

Тут y_m – це середина відповідного i -го інтервалу варіаційного ряду. Контроль: $\sum n_i = n$.

$$\sum_{i=1}^{15} n_i = 2 + 15 + 16 + 13 + 9 + 10 + 15 + 5 + 6 + 4 + 2 + 2 + 0 + 0 + 1 = 100.$$

Проаналізуємо отриманий згрупований розподіл частот вибірки для величини Y за допомогою полігону та гістограми частот.

Полігоном частот називають ламану лінію, відрізки якої з'єднують точки (x_1, n_1) , (x_2, n_2) , ..., (x_k, n_k) . Для побудови полігону

частот на осі абсцис відкладають варіанти x_i , а на осі ординат – їх частоти n_i .

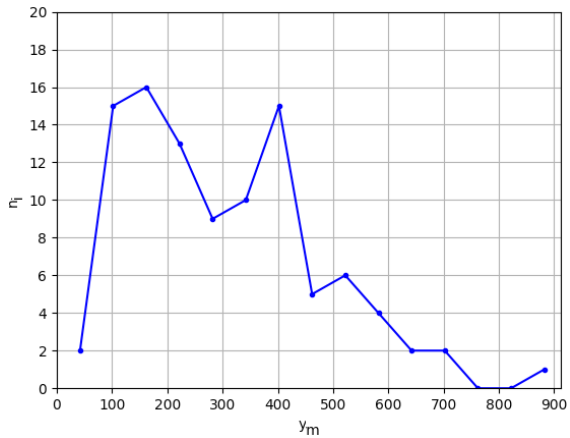


Рис. 2. Полігон частот для Y

Гістограмою частот називають фігуру, яка складається з прямокутників, основи яких дорівнюють інтервалам довжиною h , а висоти – частотам n_i .

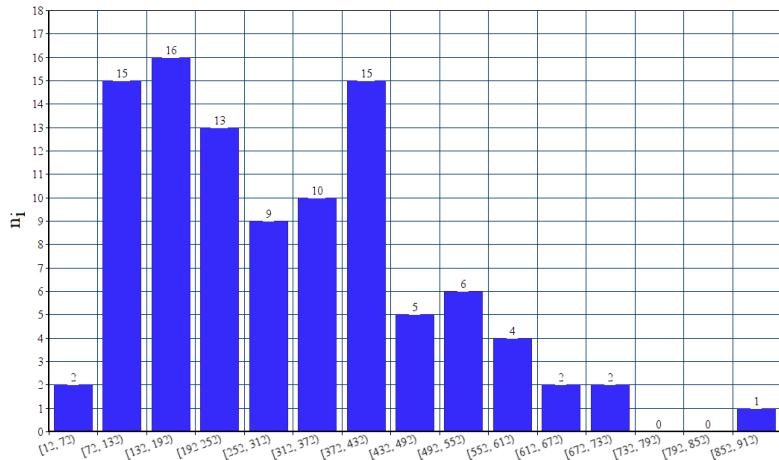


Рис. 3. Гістограма частот для Y

Побудова емпіричної функції розподілу

Функцією розподілу називають таку функцію $F(X)$, яка визначає ймовірність того, що випадкова величина X в результаті випробування набуде значення менше за x .

$$F(x) = P(X < x). \quad (5.1)$$

Геометрично цю рівність можна інтерпретувати так: $F(x)$ – ймовірність того, що випадкова величина X набуде значення, яке зображене на числовій осі точкою розташованою трохи ліворуч від значення x (див. рис. 4).

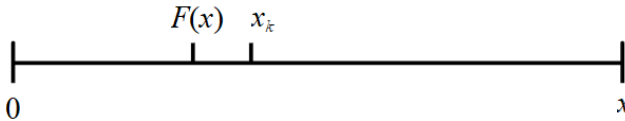


Рис. 4. Геометрична інтерпретація функції розподілу $F(x)$

Іноколи замість терміну «функція розподілу» вживається термін «інтегральна функція розподілу».

Функцією розподілу вибірки є емпірична функція розподілу.

Емпіричною функцією розподілу називають таку функцію $F^*(x)$, яка визначає для кожного значення x відносну частоту події $X < x$.

$$F^*(x) = n_i / n, \quad (5.2)$$

де n_i – число варіантів менших за x ; n – обсяг вибірки.

Властивості емпіричної функції розподілу:

- значення емпіричної функції належать відрізку $[0, 1]$;
- $F^*(x)$ – зростаюча функція;
- якщо x_1 – найменша варіанта, то $F^*(x) = 0$ при $x \leq x_1$, якщо x_k – найбільша варіанта, то $F^*(x) = 1$ при $x > x_k$.

Алгоритм побудови емпіричної функції розподілу:

- знаходження об'єму вибірки;
- визначення мінімального значення вибірки x_{min} , $F^*(x) = 0$ при $x \leq x_{min}$;

- обчислення $F^*(x)$ та визначення меж інтервалів для $x > x_{min}$;
- визначення максимального значення вибірки x_{max} , $F^*(x) = 1$ при $x > x_{max}$.

Для побудови емпіричної функції розподілу вибірки (ознаки Y), що розглядається (див. завдання до курсової роботи), скористаємось значеннями табл. 2 (y_m).

$$F^*(y) = \begin{cases} 0 & \text{при } y \leq 42, \\ 0,02 & \text{при } 42 < y \leq 102, \\ 0,17 & \text{при } 102 < y \leq 162, \\ 0,33 & \text{при } 162 < y \leq 222, \\ 0,46 & \text{при } 222 < y \leq 282, \\ 0,55 & \text{при } 282 < y \leq 342, \\ 0,65 & \text{при } 342 < y \leq 402, \\ 0,80 & \text{при } 402 < y \leq 462, \\ 0,85 & \text{при } 462 < y \leq 522, \\ 0,91 & \text{при } 522 < y \leq 582, \\ 0,95 & \text{при } 582 < y \leq 642, \\ 0,97 & \text{при } 642 < y \leq 702, \\ 0,99 & \text{при } 702 < y \leq 822, \\ 1 & \text{при } y > 882. \end{cases}$$

Графік емпіричної функції розподілу $F^*(y)$ зображено на рис. 5.

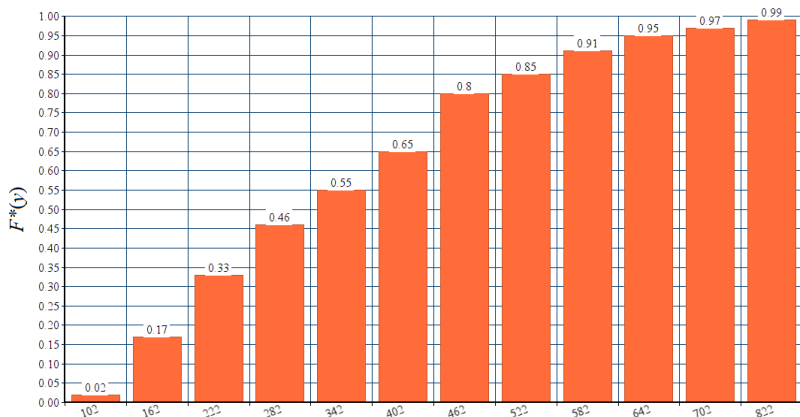


Рис. 5. Графік емпіричної функції розподілу за Y

Побудова графіка рівняння лінійної регресії

Регресія – це залежність математичного сподівання (наприклад, середнього значення) випадкової величини від однієї або кількох інших випадкових величин. Поняття регресії в деякому змісті є узагальненням функціональної залежності $y = f(x)$.

Регресійний аналіз полягає у визначенні аналітичного виразу зв'язку, в якому зміна однієї величини (результативної ознаки) обумовлено впливом однієї або кількох незалежних величин.

Один із базових методів регресійного аналізу для оцінки невідомих параметрів регресійних моделей за вибірковими даними є метод найменших квадратів (МНК, англ. LS – Least Squares).

Метод найменших квадратів – це математичний метод, який застосовується для вирішення різного роду завдань і заснований на мінімізації суми квадратів відхилень функцій від шуканих змінних.

За формою залежності розрізняють два види регресій:

- *лінійну*,
- *криволінійну*.

Лінійна регресія — це метод моделювання залежності між скалярною змінною Y та векторною (у загальному випадку) змінною X . У разі, якщо змінна X також є скаляром, регресію називають *простою*.

Лінійна регресія виражається рівнянням прямої:

$$Y = aX + b, \quad (6.1)$$

де a та b – параметри рівняння, які необхідно знайти.

Визначення значень параметрів a та b зводиться до розв'язку системи рівнянь виду:

$$\begin{cases} \left(\sum_{i=1}^n x_i^2 \right) a + \left(\sum_{i=1}^n x_i \right) b = \sum_{i=1}^n x_i y_i, \\ \left(\sum_{i=1}^n x_i \right) a + nb = \sum_{i=1}^n y_i, \end{cases} \quad (6.2)$$

де x_i та y_i – варіанти ознаки X та Y , n – об'єм вибірки.

Розв'язавши систему рівняння (6.2), отримуємо рівняння лінійної регресії Y на X . Скориставшись значеннями вибірки з табл. 1 та розв'яжемо систему (6.2). У результаті одержимо таке рівняння лінійної регресії:

$$Y = -1,1649 X + 364,2552. \quad (6.3)$$

Побудуємо графік лінійної регресії та співставимо його з фактичними значеннями вибірки (див. рис. 6).

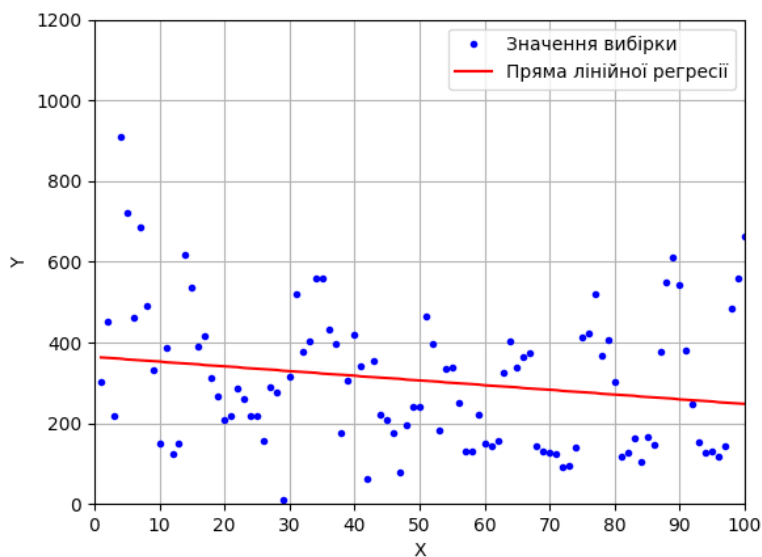


Рис. 6. Графік лінійної регресії Y на X

Побудова графіка рівняння криволінійної регресії

Криволінійний зв'язок між ознаками – це такий зв'язок, при якому рівномірним змінам першої ознаки відповідають нерівномірні зміни другої, причому ця нерівномірність має певний закономірний характер.

Іншими словами, при криволінійній формі зв'язку збільшення факторної ознаки призводить до нерівномірного збільшення (або зменшення) результуючої ознаки, тобто зростання її величини змінюється на спадання, а зменшення – збільшенням.

Якщо графік регресії $y = f(x)$ зображається кривою лінією, то таку регресію називають *криволінійною*.

Криволінійну регресію виражають за допомогою рівняння регресії 2-го, 3-го або більшого порядку. Розглянемо рівняння криволінійної регресії 2-го та 3-го порядку:

$$Y = aX^2 + bX + c, \quad (7.1)$$

$$Y = aX^3 + bX^2 + cX + d, \quad (7.2)$$

де a , b , c та d – параметри рівнянь, які необхідно знайти.

Визначення цих параметрів зводиться до розв'язку систем рівнянь виду:

$$\begin{cases} \left(\sum_{i=1}^n x_i^4 \right) a + \left(\sum_{i=1}^n x_i^3 \right) b + \left(\sum_{i=1}^n x_i^2 \right) c = \sum_{i=1}^n x_i^2 y_i, \\ \left(\sum_{i=1}^n x_i^3 \right) a + \left(\sum_{i=1}^n x_i^2 \right) b + \left(\sum_{i=1}^n x_i \right) c = \sum_{i=1}^n x_i y_i, \\ \left(\sum_{i=1}^n x_i^2 \right) a + \left(\sum_{i=1}^n x_i \right) b + n c = \sum_{i=1}^n y_i, \end{cases} \quad (7.3)$$

$$\begin{cases}
 \left(\sum_{i=1}^n x_i^6 \right) a + \left(\sum_{i=1}^n x_i^5 \right) b + \left(\sum_{i=1}^n x_i^4 \right) c + \left(\sum_{i=1}^n x_i^3 \right) d = \sum_{i=1}^n x_i^3 y_i, \\
 \left(\sum_{i=1}^n x_i^5 \right) a + \left(\sum_{i=1}^n x_i^4 \right) b + \left(\sum_{i=1}^n x_i^3 \right) c + \left(\sum_{i=1}^n x_i^2 \right) d = \sum_{i=1}^n x_i^2 y_i, \\
 \left(\sum_{i=1}^n x_i^4 \right) a + \left(\sum_{i=1}^n x_i^3 \right) b + \left(\sum_{i=1}^n x_i^2 \right) c + \left(\sum_{i=1}^n x_i \right) d = \sum_{i=1}^n x_i y_i, \\
 \left(\sum_{i=1}^n x_i^3 \right) a + \left(\sum_{i=1}^n x_i^2 \right) b + \left(\sum_{i=1}^n x_i \right) c + n d = \sum_{i=1}^n y_i.
 \end{cases} \quad (7.4)$$

Розв'язати системи (7.3) та (7.4) можна за допомогою методу Гауса або Крамера. Для значень вибірки, взятих із табл. 1, отримаємо такі рівняння криволінійної регресії 2-го та 3-го порядку та, відповідно, їх графіки (див. рис.7).

$$Y = 0,0706 X^2 - 8,2995 X + 485,5442, \quad (7.5)$$

$$Y = 0,0002 X^3 + 0,0405 X^2 - 7,077 X + 474,9996. \quad (7.6)$$

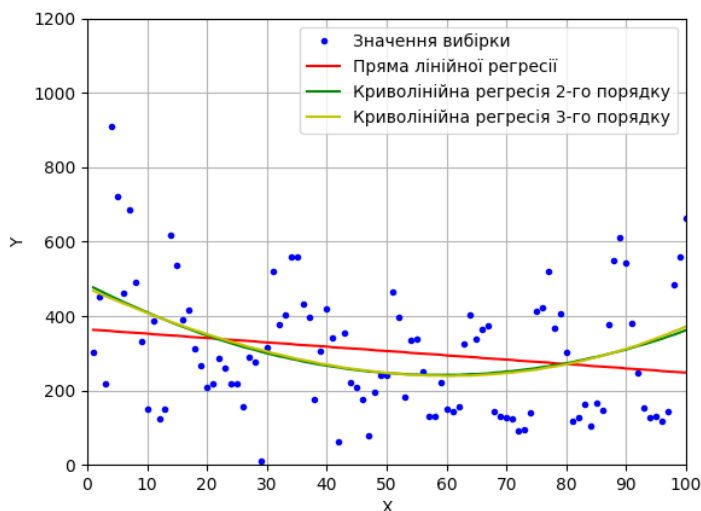


Рис. 7. Графіки криволінійної регресії 2-го та 3-го порядку Y на X

Розрахунок довірчих інтервалів

Розглянуті раніше числові характеристики є точковими оцінками, але поруч із ними щодо вибірки використовуються також й інтервальні оцінки, оскільки важливим є не тільки виконати оцінку, а й охарактеризувати (врахувати) величину можливої похибки.

Інтервальною називають таку оцінку, яка визначається двома числами – кінцями інтервалу. Інтервальні оцінки дозволяють встановити точність і надійність оцінок. Кінці інтервалів є також випадковими величинами.

Величина δ характеризує *точність оцінки*, якщо виконується нерівність:

$$|\Theta - \Theta^*| < \delta \quad (8.1)$$

де Θ^* – оцінка деякого параметра Θ генеральної сукупності.

Надійністю оцінки (довірчою ймовірністю) Θ^ по Θ* називають ймовірність γ , з якою виконується нерівність (8.1).

Найчастіше використовують надійність, яка набуває значення 0,95, 0,99 та 0,999.

Довірчим інтервалом називають інтервал $(\Theta^* - \delta, \Theta^* + \delta)$, який покриває невідомий параметр із заданою надійністю γ .

Якщо випадкова величина, що розглядається, має нормальний розподіл, але її середньоквадратичне відхилення невідоме, то можна побудувати довірчий інтервал по розподілу Стюдента з $k = n - 1$ ступенями вільності, тобто справедлива нерівність:

$$\bar{x}_B - t_\gamma \frac{s}{\sqrt{n}} < a < \bar{x}_B + t_\gamma \frac{s}{\sqrt{n}}, \quad (8.2)$$

де \bar{x}_B – вибіркове середнє, s – «виправлене» середньоквадратичне відхилення вибірки, n – обсяг вибірки,

a – математичне сподівання генеральної сукупності та t_γ – параметр значення якого залежить від γ та n і визначається за допомогою таблиці значень для розподілу Стюдента.

Зверніть увагу! При обсязі вибірки $n > 30$ розподіл Стюдента прямує до нормального. Тому, при побудові довірчого інтервалу доцільніше замість розподілу Стюдента використовувати нормальний розподіл, тобто

$$\bar{x}_B - t \frac{s}{\sqrt{n}} < a < \bar{x}_B + t \frac{s}{\sqrt{n}}, \quad (8.2^*)$$

де t – аргумент функції Лапласа $\Phi(t)$.

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{z^2}{2}} dz. \quad (8.3)$$

Значення $\Phi(t)$ табульовані. Крім того, функція Лапласа пов'язана зі надійністю оцінки (γ) наступним співвідношенням

$$\Phi(t) = \frac{\gamma}{2} \quad (8.4)$$

Знайдемо довірчі інтервали для математичного сподівання Y з надійністю $\gamma = 0,95, 0,99$ та $0,999$.

Довірчі інтервали. Таблиця 3

| Довірчий інтервал для $\gamma = 0,95$ | | |
|--|------------|----------|
| 271,9358 | $< M(Y) <$ | 338,9242 |
| Довірчий інтервал для $\gamma = 0,99$ | | |
| 261,3407 | $< M(Y) <$ | 349,5193 |
| Довірчий інтервал для $\gamma = 0,999$ | | |
| 247,3278 | $< M(Y) <$ | 363,5322 |

Виконання прогнозу

Прогноз (від грец. πρόβωσις «передбачення») – це науково обґрунтоване судження про можливі стани об'єкта в майбутньому та (або) про альтернативні шляхи та терміни їх здійснення. У вузькому сенсі – це ймовірне судження про майбутній стан об'єкта дослідження.

У нашому випадку, для того щоб, виконати прогноз, необхідно «продовжити» інтервал значень для останньої частини даних вибірки (останні 50 значень), тобто побудувати пряму (або криву) регресію, здійснивши екстраполяцію.

Екстраполяція – це наближення, тобто визначення за рядом даних значень функції інших її значень, що містяться поза цим рядом.

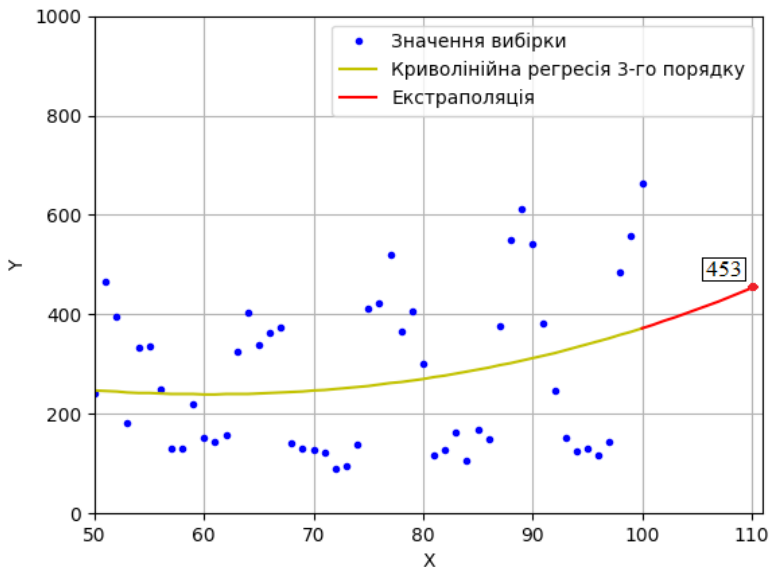


Рис. 8. Прогноз

Як видно з рис. 8., прогнозоване значення для Y станом за жовтень 2022 р. становить 453 кВт·год.

Висновки до курсової роботи

У розділі «Висновки», наведіть коротко перелік виконаних завдань. Вкажіть, чи відповідають виконані розрахунки вихідним даним і завданню до курсової роботи; якщо ні, то які розрахунки й чому не відповідають завданню.

Приклад оформлення висновків до курсової роботи

Курсова робота присвячена статистичному аналізу вимірювальних даних. У результаті написання курсової роботи були виконанні такі завдання:

1. Утворено вибірку обсягом $n = 100$ пар значень (X, Y) досліджуваної випадкової величини за допомогою одного з двох основних методів відбору (вкажіть обраний вами метод).
2. Обчислені числові характеристики вибірки:
 - вибіркове середнє;
 - вибіркова дисперсія;
 - вибіркове середньоквадратичне відхилення;
 - мода та медіана вибірки;
 - вибірковий коефіцієнт кореляції між X та Y .
3. Побудовані:
 - діаграма розсіювання;
 - гістограма та полігон частот для Y ;
 - емпірична функція розподілу за Y ;
 - графіки лінійної та криволінійної регресії Y на X
4. З надійністю $\gamma = 0,95, 0,99$ та $0,999$ обчислено довірчі інтервали для математичного сподівання генеральної сукупності досліджуваної випадкової величини.
5. Виконано прогноз.

СПИСОК ЛІТЕРАТУРИ

1. Огірко О. І., Галайко Н. В. Теорія ймовірностей та математична статистика : навч. посіб. Львів : ЛьвДУВС, 2017. 292 с.
2. Теорія ймовірностей, математична статистика та імовірнісні процеси : навч. посіб. / Ю. М. Слюсарчук, Й. Я. Хром'як, Л. Л. Джавала, В. М. Цимбал. Львів : Видавництво Львівської політехніки, 2015. 364 с.
3. Щоголев С. А. Основи теорії ймовірностей та математичної статистики : навч.-метод. посіб. Одеса : Одеський національний університет імені І. І. Мечникова, 2015. 206 с.
4. Медведев М. Г., Пащенко І. О. Теорія ймовірностей та математична статистика. Підручник. Київ : Ліра-К, 2008. 536 с.
5. Кармелюк Г. І. Теорія ймовірностей та математична статистика. Посібник з розв'язування задач. Київ : Центр навчальної літератури, 2007, 576 с.
6. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учеб. пособ. 9-е изд., стер. Москва : Высшая школа, 2004. 404 с.
7. Гмурман В. Е. Теория вероятностей и математическая статистика : учеб. пособ. 9-е изд., стер. Москва : Высшая школа, 2003. 479 с.
8. К59 Козаков О.М., Вікторовська Ю.Ю. Оформлення курсових робіт з інженерно-технічних наук : метод. рек. Чернівці : ЧНУ імені Юрія Федьковича, 2021. 56 с.

ДОДАТОК А

Розрахунок числових характеристик вибірки за допомогою Python

Утворення вибірки

```
n = int(input("Введіть обсяг вибірки n = "))
```

```
s = []
```

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

```
    s.append(float(input("Введіть "+str(i+1)+" елемент вибірки: ")))
```

Обчислення вибіркового середнього

```
sm = 0
```

```
for i in s:
```

```
    sm += i
```

```
average = sm/n
```

```
print("Вибіркове середнє:", round(average))
```

Обчислення вибіркової дисперсії

```
sm = 0
```

```
for i in s:
```

```
    sm += (i - average)**2
```

```
variance = sm/n
```

```
print("Вибіркова дисперсія:", round(variance))
```

Обчислення виправленої вибіркової дисперсії

```
corrected_variance = (n/(n-1))*variance
```

```
print("Виправлена вибіркова дисперсія:", round(corrected_variance))
```

Обчислення вибіркового середньоквадратичного відхилення

```
from math import sqrt
```

```
standard_deviation = sqrt(variance)
```

```
print("Вибіркове середньоквадратичного відхилення:",  
round(standard_deviation))
```

Обчислення виправленого середньоквадратичного відхилення

```
corrected_standard_deviation = sqrt(corrected_variance)
```

```
print("Виправлене середньоквадратичного відхилення:",  
round(corrected_standard_deviation))
```

ДОДАТОК Б

Побудова 2D-діаграми за допомогою Python

Побудова 2-D діаграми

```
import matplotlib.pyplot as plt
```

```
X = range(1, 101, 1)
```

```
x = list(X)
```

```
plt.plot(x, s, "b.", linewidth=1.5)
```

включення сітки

```
plt.grid(True)
```

задання меж для осей x та y

```
plt.xlim(1, 100)
```

```
plt.ylim(10000, 90000)
```

розбиття осі x на інтервали

```
plt.xticks(range(0,101,10))
```

позначення осей

```
plt.xlabel("X")
```

```
plt.ylabel("Y")
```

```
plt.show()
```

ДОДАТОК В

Розрахунок невизначеності вимірювань

Знаходження значення похибки для ЗВТ типу NIK 2102-02 M2B (5-60A) (див. рис. 9) в нашому випадку зводиться до оцінювання невизначеності (стандартна та розширена) результату вимірювання прямих багаторазових вимірювань за типом *A* та *B*.



Рис. 9. Зовнішній вигляд NIK 2102-02 M2B (5-60A)

В результаті вимірювального експерименту отримано $n = 100$ окремих результатів вимірювання (для Y).

Знайдемо середнє арифметичне результатів вимірювання. У нашому випадку середнє арифметичне результатів вимірювання відповідає вибіркового середньому і складає:

$$\overline{y}_B = 305 \text{ кВт} \cdot \text{год}.$$

Визначимо *стандартну невизначеність за типом A* (u_A). Оскільки результатом вимірювання обрано середнє арифметичне, то точність результату вимірювання буде характеризуватись «виправленим» середньоквадратичним відхиленням середнього результатів вимірювання, поділеним на корінь квадратний обсягу вибірки, тобто

$$u_A = \frac{s}{\sqrt{n}} = \frac{171}{\sqrt{100}} = \frac{171}{10} = 17,1 \text{ кВт} \cdot \text{год}.$$

Знайдемо *стандартну невизначеність за типом В* (u_B). Ця невизначеність обумовлена інструментальною похибкою ЗВТ.

Основна відносна похибка для NIK 2102-02 M2B (5-60A) згідно з паспортних даних становить $\delta = \pm 1,5 \%$.

Абсолютна похибка $\Delta = \delta y / 100 \% = (1,5 \times 305) / 100 \% = 4,575 \text{ кВт} \cdot \text{год}.$

Значення показів ЗВТ розподілені за нормальним законом ($n > 30$). Отже,

$$u_B = \frac{\Delta}{K_p} = \frac{4,575}{3} = 1,525 \text{ кВт} \cdot \text{год}.$$

Стандартна невизначеність результату вимірювань (оцінену за типом А та В) дорівнює:

$$u = \sqrt{u_A^2 + u_B^2} = \sqrt{17,1^2 + 1,525^2} = 17,1678 \text{ кВт} \cdot \text{год}.$$

Результат вимірювання із зазначенням стандартної невизначеності:

$$Y = 305 \text{ кВт} \cdot \text{год}; u = 17,2 \text{ кВт} \cdot \text{год}.$$

Оцінимо розширену невизначеність з рівнем довіри $P = 0,95$ та $P = 0,99$.

$$Y(P) = k \cdot u,$$

де k – фактор покриття ($k = 2$ при $P = 0,95$ та $k = 3$ при $P = 0,99$).

Результат вимірювання із зазначенням розширеної невизначеності:

$$Y = (305 \pm 34,4) \text{ кВт} \cdot \text{год}, P = 0,95;$$

$$Y = (305 \pm 51,6) \text{ кВт} \cdot \text{год}, P = 0,99.$$

Навчальне видання

Спеціальні глави вищої математики

Методичні вказівки до курсової роботи

Укладачі: Івашко Віктор Вікторович,
Фельде Христина Вікторівна.

Відповідальний редактор: Максимyak Петро Петрович