

**Міністерство освіти і науки України**

**Центральноукраїнський національний технічний університет**

**Пашинський В.А., Пашинський М.В.**

**СТАТИСТИЧНІ МЕТОДИ  
В ІНЖЕНЕРНИХ ДОСЛІДЖЕННЯХ**

**Навчальний посібник для здобувачів вищої освіти  
з інженерних спеціальностей**

**Кропивницький 2020**

Статистичні методи в інженерних дослідженнях. Навчальний посібник для здобувачів вищої освіти з інженерних спеціальностей. / В.А. Пашинський, М.В. Пашинський: – Кропивницький: ЦНТУ, 2020. – 106 с.

Навчальний посібник орієнтований на вивчення й практичне застосування статистичних методів аналізу даних при виконанні наукових та інженерних досліджень здобувачами першого (бакалаврського) та другого (магістерського) рівнів вищої освіти. Розглянуті базові поняття теорії імовірностей і математичної статистики, методи статистичної обробки вибірок випадкових величин, часто вживані закони розподілу, способи визначення їх параметрів та можливості використання в інженерних дослідженнях, методи аналітичного опису парних та багатofакторних емпіричних залежностей різних видів. Викладення теоретичного матеріалу проілюстроване прикладами розв'язання задач, характерних для наукових досліджень інженерного спрямування.

Посібник призначений для використання при вивченні окремих навчальних дисциплін здобувачами вищої освіти першого та другого освітнього рівня за спеціальностями 133 "Галузеве машинобудування", 192 "Будівництво та цивільна інженерія", а також при проведенні наукових досліджень здобувачами вищої освіти усіх рівнів, у тому числі при виконанні магістерських і дисертаційних робіт у галузі будівництва та інженерної механіки.

Схвалено Вченою радою Центральноукраїнського національного технічного університету, протокол № 9 від 27 серпня 2020 року.

Рецензенти: д.т.н., професор Пічугін С.Ф.,  
завідувач кафедри конструкцій з металу, дерева та пластмас  
Національного університету "Полтавська політехніка  
імені Юрія Кондратюка"

д.т.н., професор Клименко Є.В.,  
завідувач кафедри залізобетонних конструкцій  
та транспортних споруд  
Одеської державної академії будівництва та архітектури

© ЦНТУ, м. Кропивницький

© Пашинський В.А.

## ВСТУП

Експериментальні дослідження широко використовуються в процесі розроблення нових машин, виробів, матеріалів і будівельних конструкцій. Вони проводяться як у формі лабораторних чи виробничих активних експериментів, так і в формі пасивних експериментів, тобто спостереження за природними явищами, роботою конструкцій машин і механізмів в нормальних чи екстремальних умовах експлуатації. Для отримання повної та корисної інформації результати експериментальних досліджень потрібно аналізувати методами математичної статистики, створюючи й оперуючи імовірнісними моделями досліджених явищ. Тому знання методів аналізу експериментальних даних та вміння використовувати їх при проведенні наукових досліджень є важливим елементом наукової підготовки бакалаврів та магістрів інженерних спеціальностей.

Відомі підручники та посібники з теорії імовірностей і математичної статистики [1, 2, 4] є надто фундаментальними для першого знайомства з досить специфічними поняттями і методами статистичного аналізу даних, а тому складно сприймаються початківцями. Даний навчальний посібник містить викладені в стислій і доступній формі основні поняття теорії імовірностей і математичної статистики, а також конкретні рекомендації щодо методики статистичного аналізу експериментальних даних. Деякі із запропонованих методів аналізу можуть бути наближеними і не зовсім математично строгими, але вони є досить точними і цілком прийнятними для практичного використання в інженерних дослідженнях.

Засоби автоматизації обробки даних зорієнтовані на використання електронних таблиць Microsoft Excel, які є практично на кожному персональному комп'ютері і з якими тією чи іншою мірою уміють працювати усі здобувачі вищої освіти. Більш детальному вивченню можливостей Microsoft Excel, як інструмента для аналізу дослідних даних, сприяють приклади розв'язання характерних задач, наведені в розділі 6 та в додатку до посібника у вигляді робочої книги Microsoft Excel з ім'ям **Аналіз\_даних.xlsx**. Цей файл містить розв'язки викладених у посібнику прикладів задач статистичного аналізу і може бути основою для створення подібних розрахункових бланків.

Посібник може використовуватися при вивченні навчальних дисциплін, пов'язаних з аналізом експериментальних даних, при виконанні магістерських робіт, а також у науковій роботі студентів та аспірантів.

## 1. БАЗОВІ ПОНЯТТЯ ТЕОРІЇ ІМОВІРНОСТЕЙ І МАТЕМАТИЧНОЇ СТАТИСТИКИ

Складність і багатофакторність процесів, що відбуваються в суспільстві, природі та техніці, часто приводить до непередбачуваності їх результатів. Такі процеси, явища, події та величини називаються випадковими та аналізуються з використанням методів теорії імовірностей і математичної статистики. Нижче викладені основні поняття та широко вживані методи статистичного аналізу, застосування яких дозволяє отримувати ряд практично корисних результатів, а також може служити базою для глибшого вивчення імовірнісно-статистичних методів за літературою [1...4] та іншими джерелами.

**Випадковою** називають подію, яка в результаті проведення досліду (випробування, вимірювання, спостереження тощо) може реалізуватися або не реалізуватися. Наприклад, будь-яка побутова техніка може успішно відпрацювати гарантованих термін експлуатації або відмовити раніше, метеорологічний прогноз може підтвердитися або ні, кандидат на виборну посаду може бути обраний або не набрати достатню кількість голосів. Такого роду події неможливо передбачити абсолютно точно, але можна на підставі наявної інформації спрогнозувати можливість їх реалізації з тією чи іншою мірою впевненості.

Об'єктивною числовою мірою можливості реалізації випадкової події є її **імовірність**  $P$ , яка може приймати значення в межах від нуля для неможливої події до одиниці для достовірної події. Проміжні значення імовірності можуть вказувати на ступінь достовірності, можливість успішного виконання певної роботи, величину ризику тощо. На практиці подія з імовірністю реалізації  $P \approx 0$  вважається **практично неможливою**, а подія з імовірністю реалізації  $P \approx 1$  – **практично достовірною**. Такий підхід до класифікації подій на практично неможливі та практично достовірні називають принципом практичної впевненості.

Допустима величина відхилення імовірності від нуля чи одиниці для визнання події практично неможливою чи практично достовірною залежить від значимості та наслідків цієї події. Якщо згідно з метеопрогнозом імовірність дощу становить 0,8, цю подію доцільно вважати практично достовірною та вийти з дому з парасолькою. Така сама імовірність 0,8 успішного перельоту літаком є явно недостатньою, щоб визнати подію успішного перельоту практично достовірною та впевнено взяти квиток на цей рейс. Імовірність відмови побутового телевізора 0,05

протягом терміну гарантії є прийнятною, щоб вважати її практично неможливою, але така сама імовірність відмови 0,05 для ядерного реактора є надто великою небезпекою, нехтувати якою не можна.

Розглянемо класифікацію випадкових подій та елементарні операції з ними, які використовуються в інженерних та соціологічних дослідженнях, в теорії надійності, при оцінюванні якості продукції тощо.

**Несумісними** називають події, якщо в одному випробуванні (досліді, спостереженні тощо) може реалізуватися лише одна з них. Події, які можуть реалізуватися одночасно, називають **сумісними**.

Якщо внаслідок випробування обов'язково реалізується хоча б одна з визначеної групи випадкових подій, їх називають **повною групою подій**. Сума імовірностей реалізації повної групи випадкових подій завжди дорівнює одиниці. Якщо якісь об'єкти класифікуються за певною ознакою на декілька груп, то віднесення об'єкта до певної групи слід вважати несумісними подіями, які часто утворюють повну групу. Прикладами можуть сорти продукції (вищий, перший тощо), належність населеного пункту до однієї з областей або кліматичних територіальних зон держави.

Дві несумісні випадкові події, які утворюють повну групу, називають **протилежними подіями**. Відмова та безвідмовна робота деталі чи елемента будівельної конструкції в процесі експлуатації є протилежними подіями, а тому сума імовірностей цих подій завжди дорівнює одиниці.

Дві випадкові події називають **незалежними**, якщо реалізація однієї з них не впливає на імовірність реалізації іншої. **Залежними** називають випадкові події, якщо імовірність реалізації однієї з них залежить від появи чи відсутності іншої. Наприклад, обрив однієї зі шпильок, якими кріпиться колісний диск автомобіля, збільшує імовірність відмови сусідніх шпильок за рахунок зростання навантажень на них, але не впливає на імовірність відмови шпильок інших коліс. Тому руйнування шпильок кріплення одного колеса є залежними подіями, а руйнування шпильок кріплення різних коліс слід вважати незалежними подіями.

**Сумою** декількох випадкових подій називають подію, яка полягає в появі хоча б однієї з цих подій. Імовірність реалізації суми несумісних випадкових подій дорівнює сумі імовірностей цих подій. Наприклад, відмова хоча б одного стержня статично визначеної ферми перетворює її в механізм і таким чином призводить до відмови усієї ферми. Отже, відмова статично визначеної ферми є сумою елементарних подій, що полягають у відмові окремих стержнів.

*Добутком* декількох випадкових подій, називають подію, яка полягає в одночасній появі усіх цих подій. Імовірність реалізації добутку незалежних випадкових подій дорівнює добуткові імовірностей цих подій. Наприклад, безвідмовна робота статично визначеної ферми забезпечується лише за умови безвідмовної роботи усіх її стержнів. Тому імовірність безвідмовної роботи такої ферми дорівнює добутку імовірностей безвідмовної роботи усіх стержнів

Результатом спостережень чи дослідів часто бувають *випадкові величини*, значення яких неможливо передбачити до проведення дослідів. Випадковими величинами є розміри, маса та інші параметри усіх предметів, що нас оточують. Навіть якщо це не природні об'єкти, а речі розроблені людьми та виготовлені промисловим способом, вони не можуть бути абсолютно однаковими – все залежить лише від точності вимірювань. Так само випадковими і навіть більш мінливими є величини, що характеризують природні об'єкти, тектонічні чи атмосферні явища.

Кожен проведений дослід (вимірювання) дає одне конкретне значення вимірюваної величини, яке називається *реалізацією* випадкової величини. В результаті проведення серії дослідів чи спостережень формується набір значень випадкової величини, який називають *вибіркою*. Обсяг вибірки дорівнює кількості наявних значень випадкової величини у вибірці. *Генеральною сукупністю* називають множину всіх можливих реалізацій випадкової величини. Правильність формування вибірки та її обсяг порівняно з обсягом генеральної сукупності є ключовою умовою отримання точних та достовірних результатів статистичного аналізу.

*Головне завдання статистичного аналізу* даних полягає в розробленні імовірнісних моделей, які описують досліджувані явища з урахуванням їх випадкової природи. Такі моделі дозволяють прогнозувати імовірності бажаних чи небажаних випадкових подій, передбачати можливі значення випадкових величин, оцінювати рівень якості та надійності виробів, аналізувати доцільність інженерних та управлінських рішень. Предметом математичної статистики є також планування необхідного обсягу експерименту, оцінювання точності результатів спостережень, аналіз відповідності теорії до експерименту. Статистичні методи дозволяють встановлювати відсутність чи наявність зв'язку між випадковими величинами та отримувати залежності, які дають можливість прогнозувати значення одних величин через результати спостережень за іншими, а також розв'язувати багато інших практично важливих задач.

## 2. ВИПАДКОВІ ВЕЛИЧИНИ ТА ЇХ ХАРАКТЕРИСТИКИ

### 2.1. Основні властивості випадкових величин

*Випадковою* називається величина, значення якої неможливо передбачити до проведення досліду (вимірювання, спостереження). Це відповідає реальній природі матеріального світу, на явища якого впливає безліч різноманітних і непередбачуваних факторів.

Випадкові величини можуть бути дискретними та неперервними. *Дискретною* називають випадкову величину, яка в результаті досліду може приймати одне з ряду фіксованих значень. В техніці частіше зустрічаються *неперервні* випадкові величини, які відображають різні фізичні явища і значеннями яких є дійсні числа.

Множина значень, які може приймати випадкова величина в результаті досліду, утворює *область визначення* випадкової величини. Область визначення дискретної випадкової величини є множиною певних фіксованих значень, кількість яких може бути скінченою або нескінченою. Область визначення неперервної випадкової величини – множина дійсних чисел в певному інтервалі, обумовленому фізичними обмеженнями. Наприклад, швидкість вітру (по модулю) та технічні характеристики матеріалів (густина, міцність тощо) визначені в області додатних дійсних чисел, тобто  $0 \leq x < +\infty$ , а відносна вологість повітря має фізичні обмеження  $0 \leq x \leq 100\%$ .

Часом область визначення штучно звужується чи розширюється з метою спрощення імовірнісного опису випадкових величин. Так наприклад, температура атмосферного повітря та характеристики міцності матеріалів зазвичай описуються нормальним законом розподілу, який має нескінчену область визначення  $-\infty < x < +\infty$ , хоча фактично температура не може бути нижчою від абсолютного нуля, тобто  $-273^\circ\text{C}$ , а міцність не може приймати від'ємні значення. Така математична некоректність практично не впливає на результати розрахунків, оскільки імовірність виходу нормального розподілу за вказані фізичні обмеження реальних випадкових величин практично дорівнює нулю.

Випадкові величини характеризуються *законом розподілу*, який встановлює імовірність реалізації можливих значень випадкової величини, та *числовими характеристиками* (середнє значення, характеристики мінливості та форми розподілу тощо). Основні числові характеристики та часто вживані закони розподілу випадкових величин описані нижче.

## 2.2. Закон розподілу неперервної випадкової величини

Закон розподілу встановлює аналітичну, табличну чи графічну залежність між значеннями випадкової величини та імовірностями їх реалізації. Він повністю описує випадкову величину, дозволяє визначати її числові характеристики, виконувати різні перетворення й розв'язувати задачі, що виникають у науковій та інженерній практиці. Закон розподілу неперервної випадкової величини може задаватися у вигляді функції розподілу чи густини імовірності. У більшості випадків густина і функція розподілу виражаються аналітичними формулами, але наприклад, функція розподілу часто вживаного закону Гауса аналітичного виразу не має.

**Функція розподілу**  $F(x)$  дорівнює імовірності того, що випадкова величина  $X$  прийме значення, менше від  $x$  :

$$F(x) = P\{X < x\}, \quad (2.1)$$

де  $X$  і  $x$  – випадкова величина та її конкретне значення;

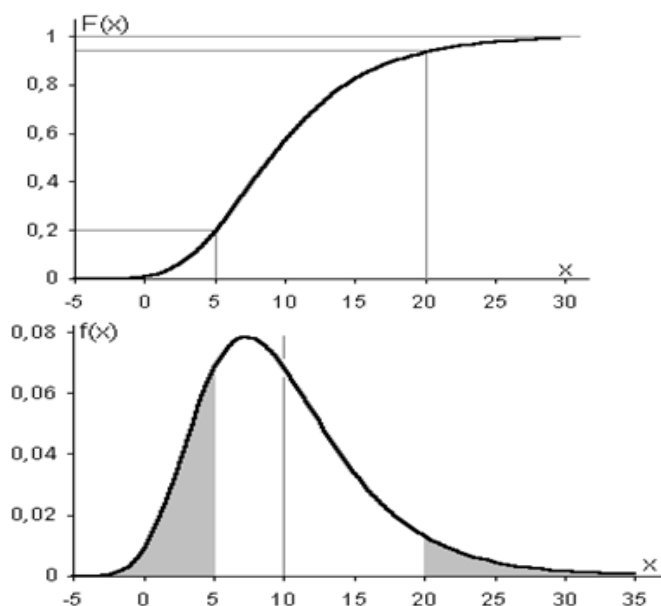
$P\{.\}$  – імовірність події, описаної у фігурних дужках.

Функція розподілу в межах області визначення випадкової величини монотонно зростає від нуля до одиниці. Якщо область визначення випадкової величини обмежена  $a \leq X < b$ , то  $F(a) = 0$ ,  $F(b) = 1$ . При нескінченній області визначення  $F(-\infty) = 0$ ,  $F(+\infty) = 1$ .

Першу похідну функції розподілу називають **густиною імовірності**, або **густиною розподілу**. Густина імовірності  $f(x)$  вказує на імовірність реалізації значень випадкової величини в околі  $x$ . Графіки густини імовірності можуть мати різноманітний характер, але всі вони повинні лежати вище від осі абсцис і задовольняти умову нормування

$$f(x) \geq 0; \quad \int_{-\infty}^{+\infty} f(x) dx = 1. \quad (2.2)$$

Характерний приклад функції та густини розподілу неперервної випадкової величини з необмеженою областю визначення наведено на рисунку 2.1. З графіка густини розподілу видно, що найімовірнішими є значення в межах приблизно від 5 до 15, а імовірність реалізації інших значень зменшується в міру віддалення від цього інтервалу.



а) функція розподілу Гумбеля  
з характеристиками:  
 $M = 10$ ;  $S = 6$   
 $F(5) = 0,195$ ;  $F(20) = 0,936$

б) густина розподілу Гумбеля  
з характеристиками:  
 $M = 10$ ;  $S = 6$   
 $P\{x < 5\} = F(5) = 0,195$   
 $P\{x > 20\} = 1 - F(20) = 0,064$

Рис. 2.1. Функція та густина розподілу неперервної випадкової величини

Густина і функція розподілу зв'язані співвідношеннями

$$f(x) = F'(x) = \frac{\partial F(x)}{\partial x}; \quad F(x) = \int_{-\infty}^x f(t) dt, \quad (2.3)$$

завдяки характеру яких функцію розподілу  $F(x)$  інколи називають *інтегральною функцією розподілу*. Зв'язок між густиною та функцією розподілу відображено на рисунку 2.1. Як слідує з формул (2.3), відмічена на графіку площа області під кривою густини розподілу при значеннях аргументу  $x < 5$  дорівнює значенню функції розподілу  $F(5)=0,195$ . Це імовірність того, що випадкова величина прийме значення менше, ніж 5. Імовірність перевищення випадковою величиною значення  $x = 20$  визначається як площа області під кривою густини розподілу при значеннях аргументу  $x > 20$  і дорівнює 0,064. Це значення може бути обчислене через функцію розподілу як  $1 - F(20) = 1 - 0,936 = 0,064$ .

Загальна формула для обчислення імовірності попадання випадкової величини в заданий інтервал  $[c, d]$  через відомі вирази функції  $F(x)$  чи густини розподілу  $f(x)$  має вигляд:

$$\text{Prob}\{c < x < d\} = F(d) - F(c) = \int_c^d f(x) dx. \quad (2.4)$$

У практичних імовірнісних розрахунках (наприклад, при визначенні забезпеченості розрахункових значень навантаження чи розрахункового опору матеріалу) часто доводиться обчислювати імовірності того, що випадкова величина прийме значення  $x < c$  або  $x > c$ . Ці імовірності

визначаються шляхом встановлення нескінчених значень однієї з меж інтегрування в формулі (2.4), що приводить до виразів:

$$\text{Prob}\{x < c\} = F(c) = \int_{-\infty}^c f(x) dx; \quad (2.5)$$

$$\text{Prob}\{x > c\} = 1 - F(c) = \int_c^{+\infty} f(x) dx. \quad (2.6)$$

Позначені символами нескінченності нижня та верхня межі інтегрування на практиці дорівнюють нижній та верхній межі області визначення конкретного закону розподілу. У багатьох випадках вони дійсно є нескінченими величинами.

Не менш важливою для практики є обернена задача, коли необхідно встановити значення випадкової величини за заданою імовірністю його перевищення (чи навпаки, неперевіщення). *Квантилем* імовірності  $P$  називають значення випадкової величини  $x$ , для якого значення функції розподілу  $F(x) = P$ . Квантиль є коренем рівняння (2.5) і залежно від виду функції чи густини розподілу визначається аналітично або чисельно. Конкретні формули для визначення квантилів найпоширеніших законів розподілу наведені в підрозділі 2.4.

Для імовірнісного подання випадкових величин використовуються закони розподілу різних видів, які відрізняються математичним обґрунтуванням, аналітичними виразами, областями визначення, виглядом густини імовірності та іншими властивостями. Поширені в технічних дослідженнях закони розподілу, методи визначення їх параметрів та використання цих законів описані нижче. При відомому виді закону розподілу його форма залежить від числових характеристик випадкової величини, основними з яких є математичне сподівання, дисперсія та стандарт, коефіцієнти варіації, асиметрії та ексцесу.

### 2.3. Числові характеристики випадкової величини

Більш простим, ніж закон розподілу, поданням випадкової величини є система числових характеристик. Найчастіше використовуються такі характеристики як математичне сподівання, дисперсія, стандарт, коефіцієнти варіації, асиметрії та ексцесу. Ці характеристики зручні для виконання певних операцій з випадковими величинами і дають уявлення про їх розмірність, середні значення та ступінь мінливості, але не можуть

бути вичерпно повним поданням випадкової величини. Випадкові величини з якісно різними законами розподілу можуть мати рівні між собою основні числові характеристики. Відмінність числових характеристик завжди вказує на відмінність законів розподілу двох випадкових величин.

**Математичне сподівання** (середнє значення) визначає положення центра розподілу випадкової величини і має розмірність самої випадкової величини. При відомій густині розподілу  $f(x)$  та області визначення  $a < x < b$  математичне сподівання може бути обчислене за формулою

$$M = \int_a^b x f(x) dx. \quad (2.7)$$

У формулі (2.7) і в наступних формулах для визначення числових характеристик межі області визначення  $[a, b]$  залежать від виду закону розподілу можуть бути довільними величинами в межах від  $-\infty$  до  $+\infty$ .

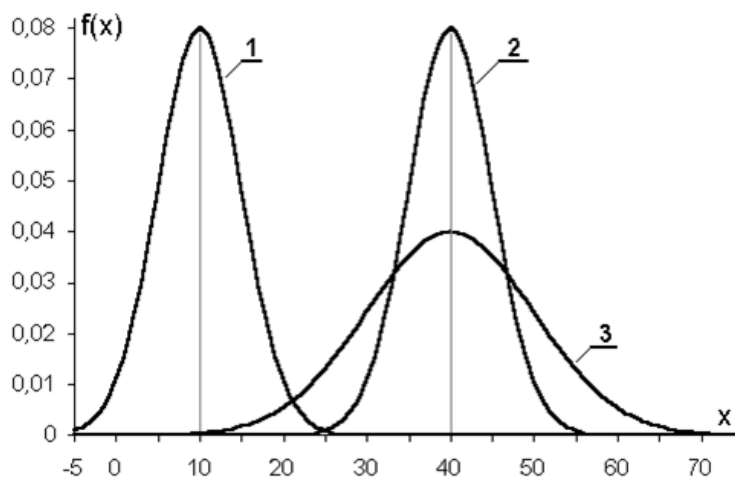
Розкид випадкової величини відносно математичного сподівання відображається **дисперсією**, яку можна обчислити за математичним сподіванням та густиною розподілу:

$$D = \int_a^b (x - M)^2 f(x) dx. \quad (2.8)$$

Дисперсія (2.8) має розмірність квадрата випадкової величини. Більш зручною для розуміння й використання характеристикою розкиду є **стандарт** (стандартне відхилення, середнє квадратичне відхилення), який має розмірність самої випадкової величини:

$$S = \sqrt{D}. \quad (2.9)$$

Смисл описаних числових характеристик проілюстровано рисунком 2.2, який відображає зміни виду густини нормального розподілу залежно від його математичного сподівання та стандарту. Математичне сподівання (2.7) визначає положення центра розподілу випадкової величини і при його зміні крива розподілу переміщається вздовж осі абсцис, не міняючи своєї форми. Дисперсія (2.8) і стандарт (2.9) характеризують міру розкиду випадкової величини відносно центра розподілу. При збільшенні стандарту крива розподілу стає ширшою, але нижчою, щоб площа під кривою густини розподілу залишалася рівною одиниці. Математичне сподівання та стандарт мають таку ж розмірність, як сама випадкова величина.



Характеристики розподілів:

1 –  $M = 10$ ,  $S = 5$

2 –  $M = 40$ ,  $S = 5$

3 –  $M = 40$ ,  $S = 10$

Рис. 2.2. Залежність густини нормального розподілу від його числових характеристик

**Коефіцієнт варіації** є відносною мірою розкиду випадкової величини та визначається за формулами

$$V = \frac{S}{M} \quad \text{або} \quad V = \frac{S}{M} \times 100\% . \quad (2.10)$$

Відповідно до використаної формули, коефіцієнт варіації є безрозмірною величиною або виражається у відсотках.

**Коефіцієнт асиметрії** відображає міру несиметричності графіка густини розподілу та визначається за формулою

$$A = \int_a^b (x - M)^3 f(x) dx / S^3 . \quad (2.11)$$

Залежність форми густини розподілу від величини коефіцієнта асиметрії відображена на рисунку 2.3.

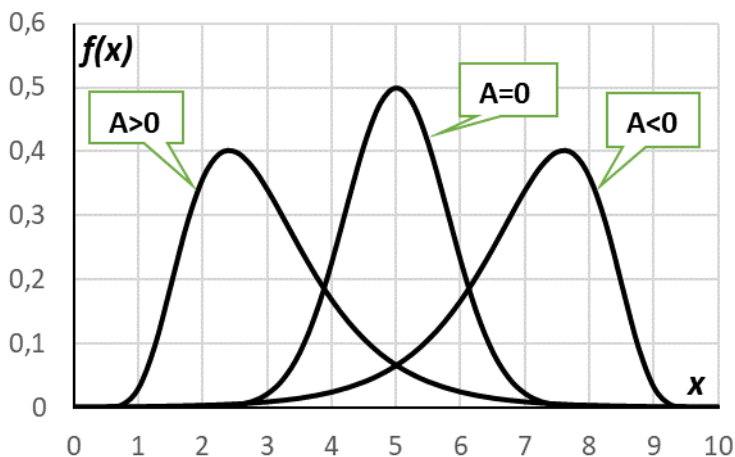


Рис. 2.3. Вплив коефіцієнта асиметрії на форму густини розподілу

З рисунка 2.3 видно, що додатне значення коефіцієнта асиметрії вказує на "розтягнутість" правої частини графіка густини розподілу, а від'ємний коефіцієнт асиметрії є наслідком більш розвинутої лівої частини густини розподілу. Нульове значення коефіцієнта асиметрії відповідає графіку густини розподілу, симетричному відносно середнього значення випадкової величини.

**Коефіцієнт ексцесу** відображає міру "гостровершинності" графіка густини розподілу та визначається за формулою

$$E = \int_a^b (x - M)^4 f(x) dx / S^4 - 3. \quad (2.12)$$

Залежність форми густини розподілу від величини коефіцієнта ексцесу відображена на рисунку 2.4.

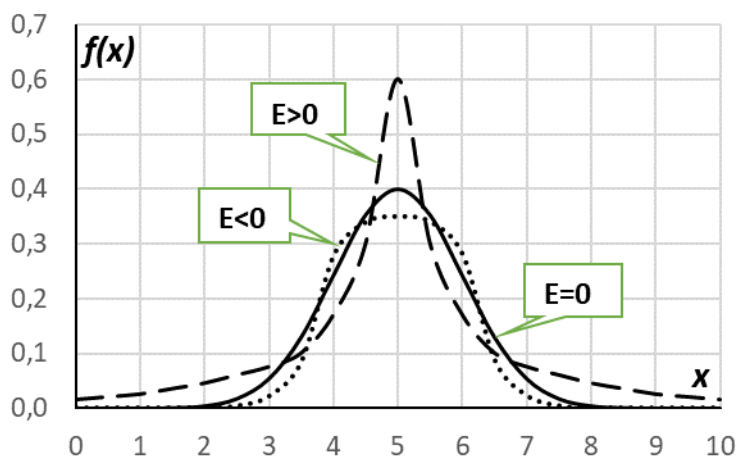


Рис. 2.4. Вплив коефіцієнта ексцесу на форму густини розподілу

З рисунка 2.4 видно, що додатне значення коефіцієнта ексцесу робить вершину кривої густини розподілу більш загостреною, а умова нормування (рівність площі під кривою густини розподілу до одиниці) виконується за рахунок зростання крайніх частин густини розподілу. При від'ємних коефіцієнтах ексцесу вершина кривої густини стає більш плоскою й широкою, а крайні частини кривої стають нижчими.

При відомих законах розподілу значення числових характеристик випадкової величини можна обчислити за загальними формулами (2.7)...(2.11), підставивши до них конкретні вирази функції чи густини розподілу. Отримані таким чином формули для визначення числових характеристик неперервних випадкових величин через відомі параметри поширених законів розподілу наведені нижче.

У деяких випадках можуть використовуватися також інші числові характеристики, смисл і методи визначення яких викладені в [1, 2, 4].

## 2.4. Поширені закони розподілу випадкових величин

Завершальним етапом статистичної обробки вибірки випадкової величини є вибір теоретичного закону розподілу, який задовільно описує проаналізовану випадкову величину і який можна використовувати для подальших досліджень. Відомі закони розподілу випадкових величин відрізняються областю визначення, формою, співвідношенням числових характеристик та іншими специфічними властивостями. Існує велика кількість різноманітних законів розподілу. Деякі з них мають певне теоретичне обґрунтування; інші застосовуються для опису окремих явищ завдяки своїм властивостям та досвіду створення імовірнісних моделей.

Вид закон розподілу обирається з урахуванням фізичної природи досліджуваної величини та вигляду гістограми розподілу, а його параметри визначаються з умови забезпечення рівності числових характеристик теоретичного розподілу та проаналізованої вибірки. Нижче описані закони розподілу, які часто використовуються в інженерних дослідженнях: рівномірний, нормальний, логнормальний, Вейбулла, експоненціальний та подвійний експоненціальний розподіл Гумбеля. Відображені їх математичні властивості, характерні особливості та методи визначення параметрів за відомими числовими характеристиками.

**2.4.1. Закон рівномірного розподілу** має простий вигляд і часто використовується для подання похибок вимірювань залежно від кроку градації шкал вимірювальних приладів. Він має функцію та густину, зображені на рисунку 2.5, і задається двома параметрами  $\alpha$  і  $\beta$ , рівними межах області визначення. Інтегральна функція та густина рівномірного розподілу виражаються формулами

$$F(x) = \frac{x - \alpha}{\beta - \alpha}; \quad f(x) = \frac{1}{\beta - \alpha}. \quad (2.13)$$

Математичне сподівання й стандарт визначаються через параметри розподілу таким чином:

$$M = \frac{(\alpha + \beta)}{2}; \quad S = \frac{(\beta - \alpha)}{\sqrt{12}}. \quad (2.14)$$

З рисунка 2.5 і виразу для густини розподілу (2.13) видно, що закон рівномірного розподілу має симетричний вигляд, а тому його коефіцієнт асиметрії завжди дорівнює нулю:  $A = 0$ .

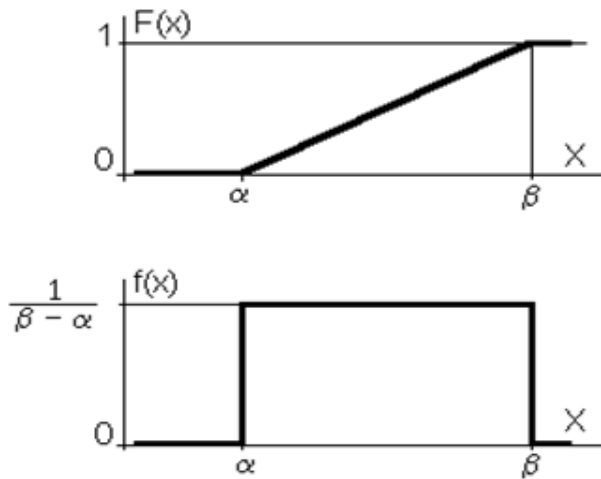


Рис. 2.5. Функція і густина рівномірного розподілу

Якщо випадкова величина має відомі фізичні обмеження, параметри  $\alpha$  і  $\beta$  приймають рівними цим обмеженням. При невідомих обмеженнях параметри можна визначати за формулами, отриманими з (2.14):

$$\alpha = M - S\sqrt{3} ; \quad \beta = M + S\sqrt{3} , \quad (2.15)$$

де  $M$  і  $S$  – оцінки математичного сподівання та стандарту розподілу.

Квантиль рівномірного розподілу імовірності  $P$  визначається через інтегральну функцію розподілу за формулою

$$x = \alpha + P \times (\beta - \alpha) . \quad (2.16)$$

Нагадаємо, що квантиль  $x$  – це значення випадкової величини, імовірність неперевіщення якого дорівнює  $P$ , а імовірність перевищення дорівнює  $1 - P$ .

**2.4.2. Нормальний закон розподілу** (розподіл Гауса) є граничним розподілом, до якого наближається сума багатьох випадкових величин. Тому він широко використовується в різноманітних галузях науки і техніки для опису випадкових величин, які формуються в результаті впливу багатьох випадкових факторів. Нормальний розподіл має нескінчену область визначення  $-\infty < x < +\infty$  і задається двома параметрами, рівними математичному сподіванню  $M$  та стандарту  $S$  випадкової величини. Зображений на рисунку 2.2 графік його густини має симетричну пагорбоподібну форму й описується формулою

$$f(x) = \frac{1}{S\sqrt{2\pi}} \exp\left[-\frac{(x-M)^2}{2S^2}\right] . \quad (2.17)$$

Інтегральна функція нормального розподілу  $F(x)$  не має аналітичного виразу, а тому визначається за спеціальними таблицями залежно від аргументу

$$t = (x - M) / S . \quad (2.18)$$

Таблиці нормального розподілу, наведені в додатку А та в літературі [1, 2, 4], дозволяють також розв'язувати обернену задачу: визначати нормований аргумент  $t$  за заданим значенням функції розподілу  $F(t)$ . Тоді квантилі нормального розподілу дорівнюють

$$x = M \pm t \times S . \quad (2.19)$$

При використанні таблиць слід пам'ятати, що  $F(-t) = 1 - F(t)$ . Для обчислень у середовищі табличного процесора Microsoft Excel можна скористатися статистичними функціями  $\text{НОРМРАСП}(x, M, S, I)$  і  $\text{НОРМОБР}(F, M, S)$ , параметри яких позначені вище. При  $I = 0$  функція  $\text{НОРМРАСП}(\dots)$  обчислює густину розподілу (2.17), а при  $I = 1$  – інтегральну функцію розподілу, відповідну загальному виразу (2.1).

**2.4.3. Логнормальний закон розподілу** (логарифмічно нормальний розподіл) описує випадкову величину, логарифм якої розподілений за нормальним законом. Він визначений в області додатних чисел  $x > 0$  та відображає граничний розподіл добутку багатьох випадкових величин. Густина логнормального розподілу, приклади графіків якої зображені на рисунку 2.6, має пагорбоподібну форму з правосторонньою асиметрією й описується виразом

$$f(x) = \frac{1}{\beta x \sqrt{2\pi}} \exp \left[ -\frac{(\ln x - \alpha)^2}{2\beta^2} \right], \quad (2.20)$$

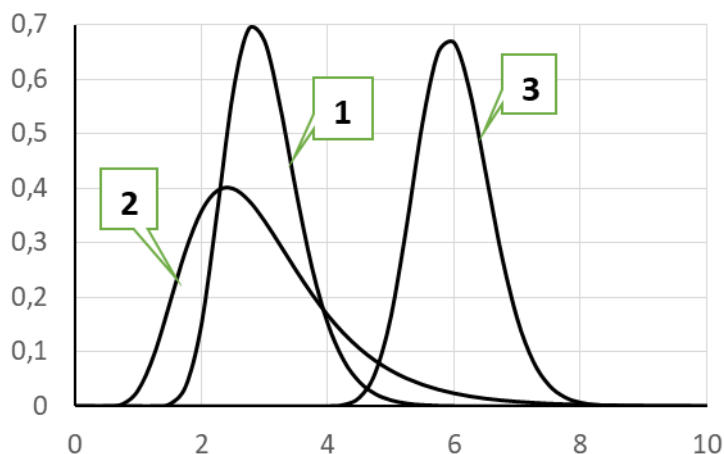
де  $\alpha$  і  $\beta$  – параметри, які визначаються через середнє значення  $M$  та коефіцієнт варіації вибірки  $V$  за формулами

$$\beta = \sqrt{\ln(V^2 + 1)} ; \quad \alpha = \ln(M) - \beta^2 / 2 . \quad (2.21)$$

Оскільки логнормальний закон не має аналітичного виразу функції розподілу, вона визначається з використанням таблиць нормального розподілу з додатка А чи з книг [1, 2, 4] через нормований аргумент

$$t = (\ln x - \alpha) / \beta , \quad (2.22)$$

або в середовищі табличного процесора Microsoft Excel з використанням статистичних функцій  $\text{НОРМРАСП}(\ln x, \alpha, \beta, I)$  і  $\text{НОРМОБР}(F, \alpha, \beta)$ .



Характеристики та параметри розподілів:

1 – M = 3	S = 0,6
α = 1,079	β = 0,198
2 – M = 3	S = 1,2
α = 1,024	β = 0,385
3 – M = 6	S = 0,6
α = 1,787	β = 0,100

Рис. 2.6. Густина логарифмічно нормального розподілу

**2.4.4. Закон розподілу Гумбеля** (подвійний експоненціальний розподіл) є граничним розподілом найбільшого елемента вибірки випадкової величини. Саме цим пояснюється його широке застосування для опису екстремальних значень різних величин: рівня води при повені, найбільших протягом місяця чи року значень кліматичних факторів тощо. Область визначення розподілу Гумбеля нескінчена  $-\infty < x < +\infty$ , а зображений на рисунку 2.1 графік густини імовірності має пагорбоподібну форму та правосторонню асиметрію з незмінним коефіцієнтом асиметрії  $A=1,14$ . Густина розподілу Гумбеля описується формулою

$$f(x) = \frac{1}{\beta} \exp \left[ \frac{\alpha - x}{\beta} - \exp \left( \frac{\alpha - x}{\beta} \right) \right], \quad (2.23)$$

а інтегральна функція розподілу Гумбеля має аналітичний вираз

$$F(x) = \exp \left[ -\exp \left( \frac{\alpha - x}{\beta} \right) \right], \quad (2.24)$$

де  $\alpha$  і  $\beta$  – параметри, які в граничному випадку (при близькому до нескінченості обсязі вибірки, з якої вибирається кожен максимум) наближено визначаються за формулами:

$$\alpha = M - 0,45 S; \quad \beta = 0,78 S. \quad (2.25)$$

Більш точно параметри розподілу Гумбеля визначаються з урахуванням обсягу  $N$  наявної вибірки максимальних значень за формулами з [3]

$$\alpha = M - S \times (0,45 + 0,34N^{-0,69}); \quad \beta = S \times (0,78 + 1,54N^{-0,75}). \quad (2.26)$$

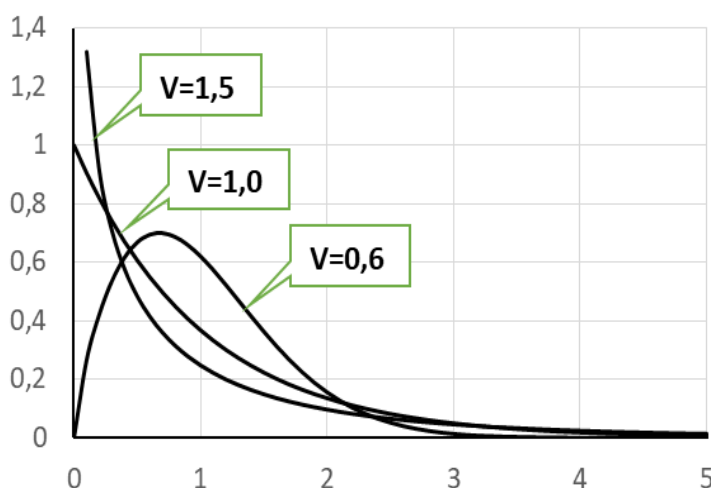
Аналітичний вираз функції розподілу (2.24) дозволяє обчислювати імовірності перевищення заданих значень випадкової величини, а також аналітично визначати квантілі, тобто значення аргументу розподілу  $x$  за заданою забезпеченістю  $P = F(x)$ . Двічі прологарифмувавши (2.24) та виконавши необхідні алгебраїчні перетворення, отримуємо вираз для квантилів розподілу Гумбеля у вигляді:

$$x = \alpha - \beta \times \ln(-\ln P) , \quad (2.27)$$

де  $P=F(x)$  – задана забезпеченість шуканого максимального значення випадкової величини.

**2.4.5. Закон розподілу Вейбулла** отримано як розподіл мінімального елемента з вибірки випадкової величини, визначеної в області додатних чисел. Він використовується в дослідженнях втомлюваності матеріалів, для опису метеорологічних факторів та ряду інших явищ.

Розподіл Вейбулла існує в області додатних дійсних чисел  $0 < x < +\infty$ , а вигляд його густини залежить від значення параметра  $\alpha$ , як показано на рисунку 2.7. Якщо  $\alpha < 1$ , розподіл має експонентоподібний характер з коефіцієнтом варіації  $V > 1$  та густиною розподілу  $f(x) = \infty$  при  $x = 0$ . При  $\alpha < 1$  отримуємо  $V > 1$ , а густина розподілу має одновершинний вигляд з переважно правосторонньою асиметрією. У випадку  $\alpha = 1$  розподіл Вейбулла вироджується в розглянутий нижче експоненціальний розподіл з коефіцієнтом варіації  $V = 1$  та коефіцієнтом асиметрії  $A = 2$ .



Числові характеристики та параметри розподілів:

$M = 1$	$V = 1,5$
$\alpha = 0,685$	$\beta = 1,193$
$M = 1$	$V = 1,0$
$\alpha = 1,000$	$\beta = 1,000$
$M = 1$	$V = 0,6$
$\alpha = 1,717$	$\beta = 0,821$

Рис. 2.7. Можливі види густини розподілу Вейбулла

Розподіл Вейбулла має аналітичні вирази густини та інтегральної функції розподілу, які мають вигляд

$$f(x) = \alpha \beta x^{\alpha-1} \exp(-\beta x^\alpha); \quad (2.28)$$

$$F(x) = 1 - \exp(-\beta x^\alpha), \quad (2.29)$$

де  $\alpha$  і  $\beta$  – параметри розподілу.

Параметр  $\alpha$  визначається як розв'язок трансцендентного рівняння

$$\Gamma\left(1 + \frac{2}{\alpha}\right) - (V^2 + 1) \times \left[\Gamma\left(1 + \frac{1}{\alpha}\right)\right]^2 = 0, \quad (2.30)$$

де  $\Gamma(\dots)$  – гамма-функція;

$V = S / M$  – коефіцієнт варіації випадкової величини.

Параметр  $\beta$  обчислюється через  $\alpha$  і математичне сподівання  $M$  випадкової величини за формулою

$$\beta = \left[ \frac{\Gamma(1 + 1/\alpha)}{M} \right]^\alpha. \quad (2.31)$$

Рівняння (2.30) можна розв'язати за допомогою функції "пошук рішення" в середовищі Microsoft Excel. Для цього необхідно записати формулу для обчислення лівої частини рівняння (2.30) в одній з комірок та визначити її як цільову функцію, яку необхідно прирівняти до нуля шляхом зміни комірки, що містить значення параметра  $\alpha$ . В якості початкового наближення для розв'язання рівняння (2.30) доцільно взяти  $\alpha \approx 1/V$ . Гамма-функції в формулах (2.30) і (2.31) обчислюються з використанням вбудованої функції Excel  $\text{ГАММА}(\dots)$ .

Коефіцієнт асиметрії  $A$  розподілу Вейбулла зростає при збільшенні коефіцієнта варіації  $V$  за залежністю, наведеною в [3]. Для орієнтовного порівняння асиметрії теоретичного розподілу Вейбулла з відповідною статистичною оцінкою можна скористатися рисунком 2.8, де наведена залежність  $A$  від  $V$ , обчислена за формулою з [3].

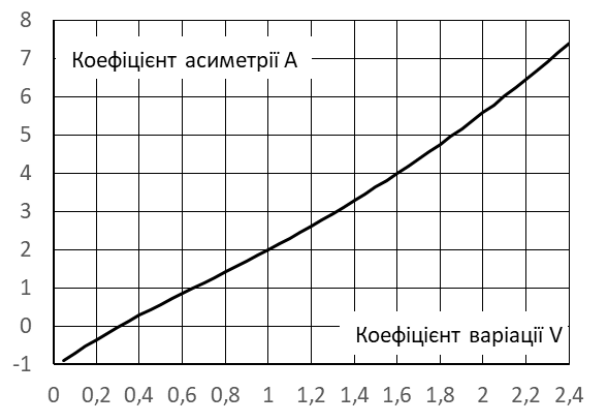


Рис. 2.8. Залежність коефіцієнта асиметрії від коефіцієнта варіації розподілу Вейбулла

Прологарифмувавши функцію розподілу (2.29), можна отримати вираз для визначення квантилів розподілу Вейбулла

$$x = \left[ -\frac{\ln(1-P)}{\beta} \right]^{1/\alpha}, \quad (2.32)$$

який дозволяє обчислювати значення випадкової величини, котре не перевищується з заданою імовірністю  $P$ .

**2.4.6. Експоненціальний закон розподілу** є окремим випадком закону розподілу Вейбулла при параметрі  $\alpha = 1$ . Він часто вживається в теорії масового обслуговування, теорії надійності та в інших галузях науки. Експоненціальний закон розподілу визначений в області додатних дійсних чисел, тобто при  $x \geq 0$ . Графік густини експоненціального розподілу зображений на рисунку 2.7, де він позначений коефіцієнтом варіації  $V=1$ . Аналітичні вирази густини та інтегральної функції експоненціального розподілу мають вигляд

$$f(x) = \beta e^{-\beta x}; \quad F(x) = 1 - e^{-\beta x}, \quad (2.33)$$

де  $\beta = 1/M = 1/S$  – параметр, обернено пропорційний математичному сподіванню та стандарту.

Характерною особливістю експоненціального розподілу є рівність математичного сподівання й стандарту  $M = S$ , а також незмінні значення коефіцієнта варіації  $V = 1$  та коефіцієнта асиметрії  $A = 2$ . Логарифмування аналітичного виразу інтегральної функції розподілу (2.33) дозволяє отримати формулу для визначення квантилів експоненціального розподілу за заданою забезпеченістю  $P$ :

$$x = -\frac{\ln(1-P)}{\beta} = -M \times \ln(1-P). \quad (2.34)$$

Друга формула (2.34) отримана шляхом підстановки наведеного вище виразу для параметра  $\beta$ . Відмітимо, що формулу (2.34) можна також отримати підстановкою до (2.32) значення  $\alpha = 1$ .

**2.4.7. Зрізані розподіли** використовуються для опису випадкових величин з фізично обмеженими областями визначення. Загалом ідея зрізаного розподілу зводиться до обмеження області визначення базового закону розподілу відповідно до реальної області визначення дослідженої випадкової величини. При цьому порушується умова нормування (2.2), що треба компенсувати шляхом коригування густини базового розподілу.

Принцип апроксимації випадкової величини зрізаним розподілом пояснюється на рисунку 2.9, де тонкою лінією показано базовий розподіл Гауса, а жирною – збудований на його основі зрізаний нормальний розподіл з областю визначення  $X > 0$ . Виділена сірим кольором площа під кривою розподілу, зрізана при  $X < 0$ , "перекачується" в зону області визначення при  $X > 0$ , яка також виділена сірим кольором. При цьому ваговий множник дорівнює  $k = 1/[1 - F(0)]$ . Як правило, густина зрізаного розподілу задається виразом типу

$$f_3(x) = k f_0(x) \quad \text{при} \quad a < x < b, \quad (2.35)$$

де  $f_0(x)$  – густина відомого базового розподілу;

$k$  – ваговий множник, який забезпечує умову нормування (2.2);

$a, b$  – межі області визначення зрізаного розподілу, одна з яких може співпадати з відповідною межею базового розподілу, у тому числі може бути нескінченною.

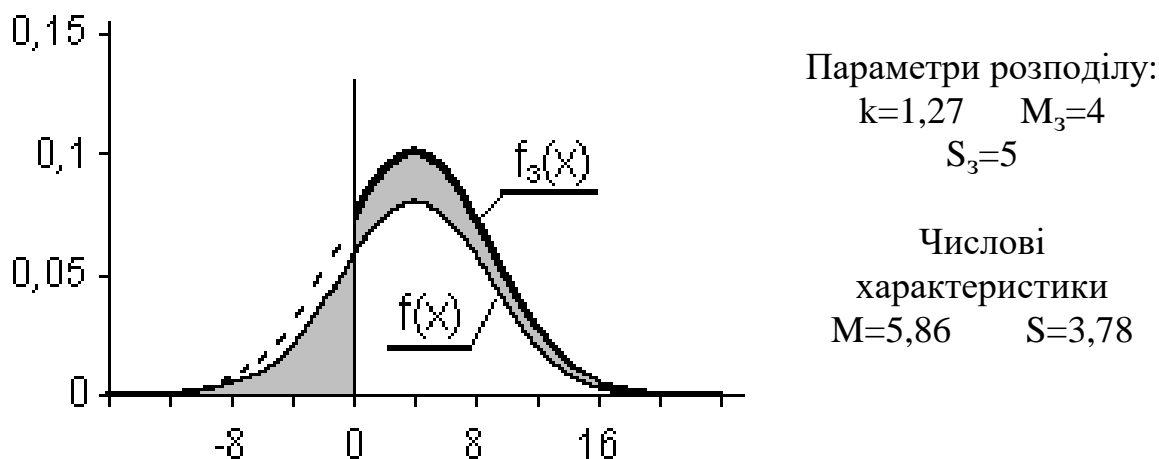


Рис. 2.9. Приклад зрізаного нормального розподілу

Проблема апроксимації зрізаним розподілом полягає в тому, що параметри  $M_3$  і  $S_3$  базового розподілу Гауса  $F_0(x)$  (чи іншого використаного базового розподілу) не дорівнюють математичному сподіванню  $M$  і стандарту  $S$  дослідженої випадкової величини, що можна побачити з рисунка 2.9. Вони повинні визначатися з умови рівності числових характеристик отриманого зрізаного розподілу до відповідних результатів статистичної обробки дослідженої вибірки. Як правило, визначення параметрів зрізаного розподілу зводиться до нетривіальної задачі чисельного розв'язання системи трансцендентних рівнянь.

## 2.5. Операції з випадковими величинами

Наявність закону розподілу та числових характеристик випадкової величини дозволяє розв'язувати низку практично важливих наукових та інженерних задач. У практиці досліджень часто виникають такі задачі:

- визначення імовірності попадання випадкової величини в заданий інтервал чи імовірності виходу за задані межі;
- встановлення мінімальних та максимальних розрахункових значень випадкової величини за заданою забезпеченістю;
- лінійні перетворення однієї випадкової величини (додавання константи або множення на константу);
- лінійні перетворення декількох випадкових величини (додавання, віднімання чи визначення лінійної комбінації випадкових величин);
- нелінійні перетворення однієї чи декількох випадкових величини (множення, ділення, інтегрування тощо).

Найчастіше доводиться оцінювати **імовірність попадання випадкової величини** до заданого інтервалу. Для цього в загальні формули (2.4), (2.5), (2.6) слід підставити конкретний вираз функції  $F(x)$  чи густини розподілу  $f(x)$  і виконати необхідні перетворення. Для законів розподілу Гумбеля, Вейбулла, експоненціального, рівномірного, а також інших розподілів, які мають аналітичний вираз  $F(x)$ , імовірність попадання до інтервалу обчислюється аналітично. Для нормального та логнормального розподілів слід скористатися таблицями з додатку А, книг [1, 2, 4] або відповідними функціями математичних пакетів, наприклад описаною вище функцією  $NORMPASP(x, M, S, I)$  з електронної таблиці Microsoft Excel. В деяких випадках обчислення за формулами (2.4), (2.5), (2.6) виконуються шляхом чисельного інтегрування густини розподілу  $f(x)$ .

Досить часто ставиться задача **нормування розрахункових значень** досліджуваних параметрів технічних об'єктів чи природних явищ. Розрахунковим звичайно вважають граничне (максимальне чи мінімальне) значення досліджуваного параметра, імовірність виходу за яке можна вважати досить малою. Імовірність протилежної події, тобто збереження параметра в бажаних межах, називають **забезпеченістю** розрахункового значення. Наприклад, характеристичний опір конструкційних матеріалів звичайно встановлюється рівним нижній межі показника міцності з забезпеченістю 0,95. Це означає, що міцність 95% випробуваних зразків перевищує значення характеристичного опору, а міцність 5% зразків відхиляється від нього в небезпечну, меншу сторону. Характеристичні та

розрахункові значення навантажень на будівельні конструкції, машини та механізми навпаки, приймаються рівними верхній межі навантаження з забезпеченістю, близькою до одиниці. Оскільки небезпечним явищем є перевищення розрахункових значень, імовірність саме цієї події має бути досить малою.

Позначимо забезпеченість розрахункового значення через  $P$ . Тоді розрахункове значення, встановлене за нижньою межею параметра, обчислюється як квантиль імовірності  $1-P$ . Розрахункове значення, встановлене за верхньою межею параметра, дорівнює квантилю імовірності  $P$ . Формули для обчислення квантилів часто вживаних законів розподілу наведені в підрозділі 2.4. Для обчислення квантилів нормального та логнормального розподілу можна скористатися функцією *НОРМОБР*( $F, M, S$ ) з електронної таблиці Microsoft Excel.

При розв'язанні науково-технічних задач часто доводиться оперувати з функціями, аргументами яких є випадкові величини. Розглянемо дві найпростіші операції, у яких бере участь випадкова величина з математичним сподіванням  $M_x$  і стандартом  $S_x$  та константа  $C$ .

При додаванні до випадкової величини  $X$  константи  $C$  числові характеристики суми дорівнюють

$$M_y = M_x + C, \quad S_y = S_x. \quad (2.36)$$

При множенні випадкової величини  $X$  на константу  $C$  отримуємо

$$M_y = C \times M_x, \quad S_y = C \times S_x. \quad (2.37)$$

Відмітимо, що віднімання константи еквівалентне додаванню тієї ж константи з протилежним знаком, а ділення на константу еквівалентне множенню на обернену величину. Додавання константи чи множення на детерміновану величину не впливає на вид закону розподілу результуючої випадкової величини, але змінює його параметри відповідно до змін математичного сподівання й стандарту за формулами (2.36) чи (2.37).

Лінійні комбінації декількох випадкових величин, а також нелінійні перетворення є більш складними операціями і часто приводять до зміни виду закону розподілу. Найпростіші методи визначення числових характеристик і законів розподілу таких функцій випадкових величин викладені в розділі 5.

### 3. СТАТИСТИЧНА ОБРОБКА ВИБІРОК ВИПАДКОВИХ ВЕЛИЧИН

#### 3.1. Загальний порядок формування і статистичної обробки вибірки випадкової величини

Завдання статистичного аналізу даних зазвичай полягає в отриманні закону розподілу випадкової величини за результатами статистичної обробки наявної вибірки. При цьому необхідно вибрати вид закону розподілу та визначити його параметри за числовими характеристиками, отриманими при статистичній обробці вибірки. Такі характеристики називають вибірковими, або оцінками числових характеристик. Методи їх обчислення викладені нижче.

Предметом статистичного аналізу є вибірка значень випадкової величини, отримана в результаті вимірювань чи спостережень. **Вибірка** – це випадковим чином обрана частина генеральної сукупності об'єктів дослідження чи даних. Під **генеральною сукупністю** розуміють повний набір усіх існуючих об'єктів чи даних. **Репрезентативність** – це властивість вибірки адекватно відображати характеристики генеральної сукупності. Для забезпечення репрезентативності вибірка повинна бути випадковою і мати достатньо великий обсяг. Найточніші результати можна отримати при аналізі усієї генеральної сукупності, але такий підхід може бути економічно недоцільним і навіть фізично неможливим. Наприклад, випробування на міцність усіх зразків виготовлених будівельних виробів дасть повні, достовірні й точні результати, але на практиці такий підхід просто не має сенсу. Тому з метою забезпечення репрезентативності вибірок використовують декілька способів відбору даних, описаних нижче.

**Випадковий відбір** забезпечується шляхом використання таблиць чи датчиків випадкових чисел, які вказують на порядкові номери об'єктів у вибірці. Наприклад, на комп'ютері за допомогою спеціальної програми генерується задана кількість порядкових номерів виробів, а потім на складі відбираються ці вироби з урахуванням порядку їх складування. Набір характеристик відібраних виробів (розміри, маса, міцність тощо) представляє собою вибірку для статистичного аналізу.

**Механічний відбір** здійснюють за алгоритмом, який забезпечує незалежність властивостей об'єктів від їх порядкового номера чи інших ідентифікаційних характеристик. Усі одиниці генеральної сукупності розташовуються в певному порядку (за зростанням або зменшенням якоїсь незалежної від мети дослідження ознаки, за алфавітом, географічним

положенням, послідовністю виготовлення тощо), а потім суто механічно через певний інтервал відбираються елементи у вибірку. Наприклад, для випробування відбирають кожний сотий виріб, випущений заводом. Випадковий вибір людей для анкетування можна здійснити шляхом систематичного відбору кожного десятого (двадцятого, сотого тощо) із загального алфавітного списку.

**Типовий відбір** полягає в тому, що генеральну сукупність спочатку ділять на типові групи за досліджуваною ознакою, а потім із кожної групи випадковим або механічним способом відбирають необхідну кількість одиниць. При цьому слід забезпечити пропорційність представництва кожної групи відповідно з їхньою чисельністю. Розчленування сукупності на типові групи дає можливість усунути вплив систематичної міжгрупової мінливості на результати аналізу вибірки. Так наприклад, при проведенні соціологічних опитувань населення забезпечується представництво всіх груп за різними ознаками (стать, вік, місце проживання, освіта тощо). Тоді випадково вибрані представники груп загалом адекватно відображають суспільну думку відповідно до структури усього суспільства.

На практиці часом використовують комбіновані методи формування вибірок, але в усіх випадках слід забезпечувати достатній обсяг вибірки та випадковість вибору окремих об'єктів дослідження. Обсяг вибірки бажано встановлювати з урахуванням точності результатів дослідження, методи прогнозування якої розроблені в математичній статистиці [1, 2, 4] та описані нижче. Бувають ситуації, коли вибірка формується незалежно від волі дослідника. Наприклад, метеорологічні спостереження за кліматом здійснюються на вже існуючій мережі метеорологічних станцій, а тому доступними є лише результати цих спостережень, а не значення кліматичних параметрів у довільно обраній географічній точці.

При будь-якому способі відбору результатом є вибірка, яка складається з  $N$  однорідних даних. Кількість даних  $N$  називають обсягом вибірки, а самі дані  $X_i$  – елементами вибірки чи реалізаціями випадкової величини, що досліджується. Подальший виклад методики проілюстровано **прикладом статистичної обробки** вибірки значень межі міцності бетону, отриманих у результаті випробувань стандартних кубів. Вибірка цих значень обсягом  $N = 20$ , виражених у мегапаскалях, має вигляд:

20,3	18,0	23,0	14,8	22,4	19,8	24,4	18,2	20,4	18,4
17,7	19,3	22,0	19,4	20,7	20,1	18,9	17,3	16,6	19,9

Перегляд наявних даних показує, що межа міцності змінюється в межах від  $X_{min} = 14,8$  МПа до  $X_{max} = 24,4$  МПа.

Статистична обробка вибірок випадкових величин звичайно виконується на персональних комп'ютерах із використанням відомих програмних комплексів. Вибірки незначного обсягу можна також обробити за допомогою калькулятора з умонтованим режимом статистики. Існує велика кількість різноманітних програмних комплексів для статистичного аналізу даних, які відрізняються набором операцій, можливостями подання результатів, зручністю інтерфейсу та доступністю для використання. Нижче рекомендовані способи статистичної обробки даних у середовищі табличного процесора Microsoft Excel, який має достатньо широкий набір вбудованих статистичних функцій, доступний практично на кожному персональному комп'ютері та принаймні в загальних рисах відомий переважній більшості користувачів.

Статистична обробка вибірок випадкових величин включає ряд етапів, перелічених нижче і детально описаних у наступних підрозділах.

1. Формування вибірки даних на машинних носіях інформації та її попередній аналіз на предмет наявності очевидно грубих помилок (похибок запису результатів спостережень чи занесення в пам'ять комп'ютера). Для цього можна відсортувати вибірку за зростанням чи зменшенням даних, визначити й оцінити мінімальне та максимальне значення. Ці операції легко виконати в середовищі Microsoft Excel.
2. Побудова гістограми розподілу, яка є статистичним еквівалентом густини імовірності випадкової величини і графічно відображає імовірності реалізації різних значень випадкової величини.
3. Обчислення числових характеристик (розмах вибірки, середнє значення, стандарт, коефіцієнти варіації та асиметрії).
4. Перевірка елементів вибірки, які різко відхиляються від основної маси даних (артефактів), на належність до вибірки. Така перевірка здійснюється з використанням спеціальних статистичних критеріїв, описаних в літературі з математичної статистики. Перевірка здійснюється послідовно, по одному елементу, який найбільше відхиляється від центра вибірки. Якщо перевірений елемент визнаний грубою помилкою і відкинутий, необхідно заново визначити числові характеристики і відкоригувати гістограму розподілу. Рішення щодо визнання якогось елемента вибірки помилковим слід приймати дуже обережно, особливо у випадку істотно асиметричних розподілів.

5. Вибір виду закону розподілу для імовірнісного опису досліджуваної випадкової величини, визначення його параметрів, побудова кривої густини імовірності на гістограмі та візуальний аналіз їх відповідності.
6. Перевірка обраного закону розподілу на відповідність до дослідних даних з використанням критеріїв узгодженості за методикою, описаною нижче. Якщо обраний закон розподілу не суперечить дослідним даним, його можна використовувати для імовірнісного опису дослідженої випадкової величини та оперування з нею при розв'язанні наукових чи інженерних задач.

Результатом статистичної обробки є числові характеристики та закон розподілу випадкової величини, які дозволяють оцінювати імовірності попадання дослідженої величини в певні межі, встановлювати розрахункові значення за заданою забезпеченістю, а також розв'язувати інші наукові та інженерні задачі імовірнісного характеру.

### 3.2. Побудова гістограми розподілу вибірки

Гістограма розподілу є статистичним еквівалентом густини розподілу випадкової величини. Побудова гістограми розподілу розпочинається з визначення розмаху вибірки, тобто найменшого  $X_{min}$  і найбільшого  $X_{max}$  зі спостережених значень. Отриману таким чином область визначення випадкової величини розбивають на інтервали, кількість яких  $L$  вибирають залежно від обсягу вибірки  $N$  згідно з рекомендацією таблиці 3.1.

Таблиця 3.1 – Рекомендована кількість інтервалів гістограми розподілу

$N =$	10 – 40	40 – 100	100 – 300	300 – 1000	1000 – 5000
$L =$	4 – 6	5 – 8	6 – 10	8 – 15	10 – 25

Межі інтервалів  $C_j$ ,  $j=0, \dots, L$  встановлюють, виходячи з довжини інтервалу, обчисленої за формулою

$$d = (X_{max} - X_{min}) / L \quad (3.1)$$

та округленою до значення, зручного для подальшого використання. Для охоплення усіх даних нижня межа першого інтервалу  $C_0$  округлюється в менший бік, а верхня межа останнього  $C_L$  – у більший. Округлення довжини інтервалу  $d$  може привести до коригування їх кількості  $L$ .

Побудова гістограми розподілу зводиться до підрахунку кількостей даних  $N_j$ , які належать до кожного з вибраних інтервалів. Це робиться шляхом перегляду вибірки вручну чи за допомогою комп'ютера. Крайні інтервали, які містять мало даних, рекомендується об'єднати. Довжина об'єданого інтервалу дорівнює сумі довжин, а кількість даних – сумі кількостей даних для інтервалів, що об'єднуються. Бажано, щоб у кожному інтервалі було не менше ніж 3...5 даних. Об'єднання інтервалів приводить до згладжування гістограми розподілу і дозволяє відмежуватися від випадкових коливань частот попадання до інтервалів при невеликих обсягах вибірок.

Імовірності попадання до кожного з інтервалів (частоти даних)  $P_j$  визначають діленням відповідних кількостей даних  $N_j$  на обсяг вибірки  $N$ :

$$P_j = N_j / N. \quad (3.2)$$

Густина імовірності в кожному інтервалі є результатом ділення частоти  $P_j$  на довжину даного інтервалу  $d_j = C_j - C_{j-1}$ :

$$f_j = \frac{P_j}{d_j} = \frac{P_j}{(C_j - C_{j-1})}. \quad (3.3)$$

Виконуючи обчислення за формулою (3.3), слід пам'ятати, що різні інтервали можуть мати різну довжину. Обчислені значення густини розподілу (3.3) дозволяють представити гістограму розподілу у вигляді графіка, приклад якого показано на рисунку 3.1.

Для побудови гістограми розподілу в середовищі Microsoft Excel рекомендується використовувати функцію *ЧАСТОТА(список,  $C_j$ )*, яка визначає кількість даних в області „список”, менших або рівних заданій межі  $j$ -того інтервалу  $C_j$ . Аргумент "список" визначає область робочої сторінки, в комірки якої занесені елементи вибірки. Для визначення списку слід просто виділити необхідну область курсором миші під час уведення функції через меню „ВСТАВКА, ФУНКЦІЯ”. Оскільки статистичні функції Microsoft Excel ігнорують пусті комірки, область робочого листа "список" може бути більшою за наявну вибірку. Це дозволяє створювати бланки, у які досить занести дані вибірки, щоб виконати їх статистичну обробку. Відмітимо, що в різних версіях Microsoft Office імена статистичних та інших функцій Microsoft Excel можуть дещо відрізнятись. Ці імена є досить близькими для розпізнання, а призначення

функцій можна уточнити за коротким описом, який надається при виборі функції.

На графічному зображенні гістограми розподілу вздовж осі абсцис відкладаються межі інтервалів, а вздовж осі ординат – значення густини імовірності (3.3), як це показано на рисунку 3.1. Побудована таким чином гістограма розподілу представляє собою сукупність прямокутників, основи яких відповідають межах інтервалів, а висоти дорівнюють густині розподілу для відповідного інтервалу. Площа кожного з прямокутників дорівнює імовірності попадання випадкової величини до відповідного інтервалу (3.2). Загальний вигляд гістограми відображає форму графіка густини імовірності дослідженої випадкової величини.

Викладена методика побудови гістограми розподілу проілюстрована **прикладом обробки** в середовищі Microsoft Excel вибірки значень межі міцності бетону, наведеної в попередньому підрозділі.

Виходячи з обсягу вибірки  $N = 20$  значень, за рекомендацією таблиці 3.1 приймаємо  $L = 5$  інтервалів гістограми розподілу. При розмаху вибірки від 14,8 МПа до 24,4 МПа за формулою (3.1) обчислюємо рекомендовану довжину інтервалу  $(24,4 - 14,8) : 5 = 1,92$  МПа. Округлення довжини інтервалу до 2 МПа приводить до 6 інтервалів у межах від 14 МПа до 26 МПа. Значення верхніх меж усіх інтервалів  $C_j$  наведені в таблиці 3.2.

Наступні стовпчики таблиці 3.2 містять:

$N_j$  – кількість даних в інтервалі, підрахована функцією *ЧАСТОТА(...)*;

$P_j$  – імовірність попадання випадкової величини до інтервалу за (3.2);

$f_j$  – густина імовірності в інтервалі за (3.3);

$f_{jt}$  – теоретичні значення густини імовірності нормального закону розподілу (2.17), методика обчислення яких описана нижче.

Таблиця 3.2 – Побудова гістограми розподілу межі міцності бетону

Номер інтервалу $J$	Верхня межа інтервалу $C_j$	$N_j$	$P_j$	$f_j$	$f_{jt}$
0	14				
1	16	1	0,050	0,025	0,0231
2	18	4	0,200	0,100	0,0921
3	20	7	0,350	0,175	0,1697
4	22	5	0,250	0,125	0,1444
5	24	2	0,100	0,050	0,0567
6	26	1	0,050	0,025	0,0103
С у м и :		20	1,000		

На рисунку 3.1 показана гістограма розподілу, побудована за наведеними в таблиці 3.2 значеннями верхніх меж інтервалів  $C_j$  та відповідними значеннями густини розподілу  $f_j$ .

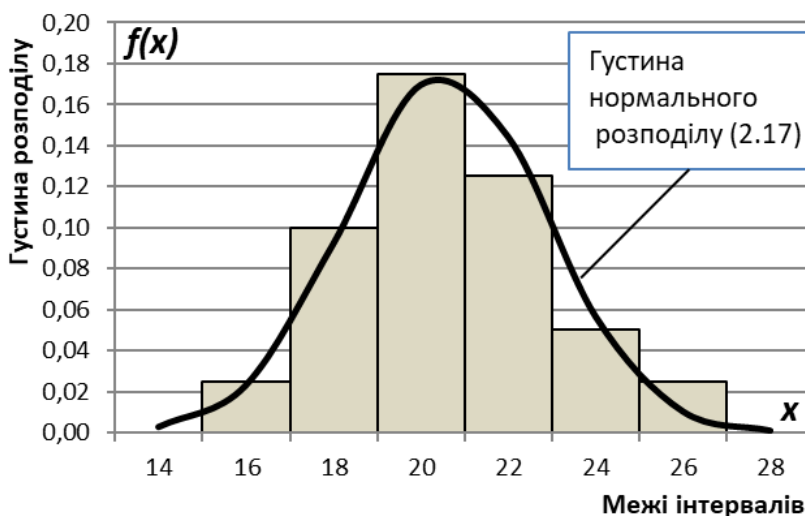


Рис. 3.1. Гістограма розподілу межі міцності бетону

Уздовж осі абсцис відкладені межі інтервалів з таблиці 3.1, а уздовж осі ординат – відповідні їм значення густини розподілу  $f_j$  з тієї ж таблиці. Підписи верхніх меж напроти середини інтервалів є особливістю графічної системи Microsoft Excel, яка при зображенні гістограм вважає дані осі абсцис іменами, а не числами, та вказує їх напроти середин інтервалів.

З рисунка видно, що збудована гістограма розподілу межі міцності бетону має досить симетричний пагорбоподібний вигляд. Зображена на гістограмі крива густини нормального розподілу (2.17), пояснення відносно якої будуть наведені нижче, візуально відповідає гістограмі.

### 3.3. Визначення числових характеристик вибірки

Числові характеристики вибірки неперервної випадкової величини можна визначити вручну за наведеними нижче формулами, або за допомогою комп'ютера з використанням математичних чи статистичних пакетів програм, які реалізують вказані формули. Нижче наведені необхідні формули та практичні рекомендації з визначення числових характеристик вибірок випадкових величин у середовищі табличного процесора Microsoft Excel, доступного практично на кожному персональному комп'ютері.

Як і в попередніх підрозділах, обсяг вибірки позначено через  $N$ , а кожен її елемент (одержане в результаті окремого вимірювання значення випадкової величини) через  $X_i$ . Поряд з розрахунковими формулами, нижче наведені рекомендації з використання вбудованих статистичних функцій Microsoft Excel. Найважчий в описі цих функцій аргумент „список” визначає область робочого аркуша, в комірці якої занесені елементи вибірки. Під час уведення функції через меню „ВСТАВКА, ФУНКЦІЯ” для визначення списку слід просто виділити необхідну область курсором миші. Раніше було сказано, що виділена область робочого аркуша може бути більшою за наявну вибірку, оскільки пусті комірки статистичними функціями ігноруються. Імена описаних нижче функцій в різних версіях Microsoft Excel можуть дещо відрізнятись, але їх призначення легко уточнити за коротким описом, який надається при виборі функції.

**Розмах вибірки** задається найменшим  $X_{min}$  і найбільшим  $X_{max}$  спостереженими значеннями, для визначення яких в Excel використовують функції  $МИН(список)$  та  $МАКС(список)$ .

**Середнє значення** (оцінка математичного сподівання), яке визначає положення центра розподілу випадкової величини, дорівнює

$$M = \frac{1}{N} \sum_{i=1}^N X_i . \quad (3.4)$$

В Excel обчислення середнього реалізоване функцією  $СРЗНАЧ(список)$ .

**Стандарт** (стандартне відхилення, середнє квадратичне відхилення) характеризує міру розкиду випадкової величини відносно центра розподілу та обчислюється за формулою

$$S = \sqrt{\frac{\sum_{i=1}^N (X_i - M)^2}{N - 1}} . \quad (3.5)$$

В Excel стандарт обчислюється функцією  $СТАНДОТКЛОН(список)$  або  $СТАНДОТКЛОН.В(список)$ . Функція  $СТАНДОТКЛОН.Г(список)$  реалізує обчислення не за вибіркою, а за генеральною сукупністю і відрізняється величиною  $N$  замість  $N - 1$  в знаменнику формули (3.5).

**Коефіцієнт варіації** є відносною мірою розкиду випадкової величини і визначається за формулою

$$V = S/M , \quad (3.6)$$

яку потрібно записати в необхідній комірці робочого листа Excel з посиланнями на комірки, де обчислені значення  $M$  і  $S$ .

**Коефіцієнт асиметрії** є мірою несиметричності розподілу відносно середнього значення й обчислюється за формулою

$$A = \frac{N}{(N-1)(N-2)} \times \frac{\sum_{i=1}^N (X_i - M)^3}{S^3}, \quad (3.7)$$

або в середовищі Excel з використанням функції *СКОС(список)*.

Для розпочатого раніше **прикладу статистичної обробки** з використанням описаних вище статистичних функцій Microsoft Excel отримані такі числові характеристики:

розмах вибірки	$X_{min} = 14,8$ МПа,	$X_{max} = 24,4$ МПа;
середнє значення	$M = 19,58$ МПа;	
стандарт	$S = 2,28$ МПа;	
коефіцієнт варіації	$V = 0,116 = 11,6\%$ ;	
коефіцієнт асиметрії	$A = 0,139$ .	

Таким чином, середнє значення границі міцності  $M = 19,58$  МПа практично відповідає марці бетону М 200, а коефіцієнт варіації  $V = 11,6\%$  є навіть нижчим від нормативного значення для важкого бетону  $V = 13,5\%$ . Невелике значення коефіцієнта асиметрії  $A = 0,139$  вказує на зовсім незначну правосторонню асиметрію, тому отриманий розподіл можна вважати практично симетричним.

Бувають ситуації, коли випадкова величина задана гістограмою розподілу, а її вибірка є недоступною. Наприклад, у такому узагальненому вигляді часто публікуються результати метеорологічних спостережень в довідниках по клімату та інших спеціалізованих виданнях. У таких випадках числові характеристики визначають не шляхом обробки вибірки даних за формулами (3.4)...(3.7), а за наявною гістограмою розподілу.

Нехай гістограма задана межами інтервалів  $C_j$  та частотами  $P_j$ , тобто імовірностями попадання випадкової величини до цих інтервалів. Першою операцією є визначення середин усіх інтервалів за формулою

$$X_j = (C_j - C_{j-1})/2. \quad (3.8)$$

Середнє значення, стандарт і коефіцієнт асиметрії дорівнюють

$$M = \sum_{j=1}^L P_j X_j; \quad S = \sqrt{\sum_{j=1}^L P_j (X_j - M)^2}, \quad A = \sum_{j=1}^L P_j (X_j - M)^3 / S^3, \quad (3.9)$$

а коефіцієнт варіації обчислюється за формулою (3.6).

Обчислення за формулами (3.8) і (3.9) реалізуються в середовищі Excel у вигляді таблиці з такими колонками: межі інтервалів  $C_j$ , середини інтервалів  $X_j$ , частоти  $P_j$ , добуток  $P_j X_j$ , добуток  $P_j (X_j - M)^2$ .

За формулами (3.9) визначаються також числові характеристики дискретної випадкової величини, заданої набором її можливих значень  $X_j$  та частот (імовірностей) їх реалізації  $P_j$ .

З метою порівняння в умовах розглянутого *прикладу* за даними таблиці 3.2 та формулами (3.8) і (3.9) визначені числові характеристики межі міцності бетону, які отримані рівними:

середнє значення	$M = 19,60$ МПа;
стандарт	$S = 2,37$ МПа;
коефіцієнт варіації	$V = 0,121 = 12,1\%$ .

Порівняння з наведеними вище точними результатами обчислення за вибіркою показало, що розрахунок за формулами (3.9) дає результати, досить близькі до формул (3.4) і (3.5). Відмінності обумовлені невеликим обсягом вибірки, за якою була побудована гістограма розподілу.

### 3.4. Точність оцінювання числових характеристик

Числові характеристики випадкової величини обчислюються згідно з викладеною вище методикою за вибіркою, сформованою шляхом випадкового відбору з генеральної сукупності. Якби до вибірки були включені інші реалізації випадкової величини, результати визначення числових характеристик були б дещо іншими. Отже, оцінки математичного сподівання (3.4), стандарту (3.5) та усіх інших числових характеристик також є випадковими величинами з певними законами розподілу.

Математичні сподівання оцінок середнього значення, стандарту й дисперсії вибірки завжди дорівнюють самим оцінкам. Нагадаємо, що оцінки середнього значення й стандарту обчислюються за формулами (3.4) і (3.5), а дисперсія дорівнює квадрату оцінки стандарту (3.5), що витікає з формул (2.8) і (2.9). Випадковий розкид оцінок математичного сподівання й дисперсії відображається стандартами цих оцінок. Згідно з [2], стандарт оцінки математичного сподівання випадкової величини з розподілом, близьким до нормального, у можна наближено визначити за формулою

$$S_M = S / \sqrt{N} , \quad (3.10)$$

а стандарт оцінки дисперсії – за формулою

$$S_D = \sqrt{\frac{2}{N-1}} \times D, \quad (3.11)$$

де  $S$ ,  $D$ ,  $N$  – стандарт, дисперсія та обсяг вибірки, яка оброблялася.

Точність оцінювання статистичних характеристик аналізується шляхом встановлення інтервальних оцінок за методом довірчих інтервалів. Довірчим називають інтервал, який з певною імовірністю покриває досліджену величину. Можна стверджувати, що з заданою довірчою імовірністю статистична характеристика не виходить за межі довірчого інтервалу. Довірчу імовірність називають також надійністю оцінювання статистичної характеристики (математичного сподівання, стандарту чи дисперсії). Точне визначення довірчих інтервалів для статистичних характеристик випадкових величин є досить складною задачею, розв'язки якої для деяких випадків наведені в [2]. Для випадкових величин з нормальним або близьким до нормального розподілом цілком задовільні результати дають рекомендовані [2] наближені методи оцінювання довірчих інтервалів для математичного сподівання та дисперсії.

Виходячи з нормального розподілу оцінки (3.10), довірчий інтервал для математичного сподівання можна визначити таким чином:

$$\left\{ M_{min} = M - \frac{t \times S}{\sqrt{N}}, \quad M_{max} = M + \frac{t \times S}{\sqrt{N}} \right\}, \quad (3.12)$$

де  $M_{min}$  і  $M_{max}$  – нижня і верхня межа довірчого інтервалу;

$t$  – аргумент нормованого нормального розподілу,

що відповідає заданій довірчій імовірності.

Величина  $t$  визначається за таблицями нормального розподілу, наведеними в додатку А та в підручниках [1, 2, 4], виходячи з двосторонньої довірчої імовірності. Величини  $t$  для часто вживаних значень довірчої імовірності наведені в наступній таблиці:

Двостороння довірна імовірність	0,8	0,9	0,95	0,98	0,99
Аргумент нормального розподілу $t$	1,28	1,64	1,96	2,33	2,58

Виходячи з величини довірчого інтервалу (3.12), можна визначити імовірну відносну похибку оцінки математичного сподівання, яка не перевищується з заданою довірчою імовірністю

$$\Delta_M = \frac{t \times S_M}{M} \times 100\% = \frac{t \times V}{\sqrt{N}}, \quad (3.13)$$

де  $V$  – коефіцієнт варіації вибірки у відсотках.

Підставивши коефіцієнт варіації вибірки у відсотках, з (3.13) отримуємо обсяг вибірки, необхідний для оцінювання середнього значення досліджуваної випадкової величини з відносною похибкою  $\Delta_M\%$ :

$$N \geq \frac{t^2 \times V^2}{\Delta_M^2}, \quad (3.14)$$

Аналогічно математичному сподіванню, за [2] отримано вирази для наближеного оцінювання нижньої та верхньої межі довірчого інтервалу для дисперсії вибірки

$$\left\{ D_{min} = D - t \times D \sqrt{\frac{2}{N-1}}, \quad D_{max} = D + t \times D \sqrt{\frac{2}{N-1}} \right\}, \quad (3.15)$$

де  $D$  – оцінка дисперсії вибірки, статистична обробка якої виконується;

$t$  – аргумент нормованого нормального розподілу, що відповідає заданій довірчій імовірності з наведеної вище таблиці чи з [1, 2, 4].

Відносна похибка оцінювання дисперсії за вибіркою обсягу  $N$  оцінюється за формулою

$$\Delta_D = t \times \sqrt{\frac{2}{N-1}} \times 100\%, \quad (3.16)$$

а мінімально необхідний обсяг вибірки для оцінювання дисперсії випадкової величини з відносною похибкою  $\Delta_D\%$  дорівнює:

$$N \geq \frac{20000 t^2}{\Delta_D^2} + 1. \quad (3.17)$$

З урахуванням співвідношення (2.9) між дисперсією та стандартом вирази для наближеного оцінювання нижньої та верхньої межі довірчого інтервалу для стандарту вибірки можна записати у вигляді:

$$\left\{ S_{min} = S \sqrt{1 - t \times \sqrt{\frac{2}{N-1}}}, \quad S_{max} = S \sqrt{1 + t \times \sqrt{\frac{2}{N-1}}} \right\}, \quad (3.18)$$

Несиметричний довірчий інтервал для стандарту (3.18) не дозволяє безпосередньо отримати вирази для відносної похибки його оцінювання та визначення необхідного обсягу вибірки. Для наближеного розв'язання цих задач можна скористатися формулами (3.16) і (3.17) для дисперсії з урахуванням того, що унаслідок квадратичного перетворення (2.9) відносна похибка оцінювання стандарту буде приблизно удвічі меншою, ніж для дисперсії.

Формули (3.14) і (3.17) використовуються для визначення необхідної кількості спостережень при плануванні експериментальних досліджень. При виконанні технічних досліджень аргумент нормального розподілу зазвичай приймають рівним  $t = 1,64$  або  $t = 1,96$ , що згідно з наведеною вище таблицею відповідає довірчій імовірності (надійності оцінки) 0,9 чи 0,95. Невідомий до початку експерименту коефіцієнт варіації вибірки приймають за даними раніше виконаних аналогічних досліджень. Так наприклад, добре відомими є коефіцієнти варіації показників міцності та інших технічних характеристик конструкційних матеріалів, а також навантажень, діючих на будівельні конструкції.

За описаною методикою визначимо показники точності оцінювання статистичних характеристик вибірки значень межі міцності бетону в умовах розпочатого раніше *прикладу статистичної обробки*. Обчислення за формулами (3.10)...(3.18) з урахуванням наведених у попередньому підрозділі оцінок числових характеристик вибірки та довірчої імовірності 0,9 з  $t = 1,64$  дають такі результати:

Показники точності	Формули	Результати
Стандарт оцінки математичного сподівання	(3.10)	0,51 МПа
Стандарт оцінки дисперсії	(3.11)	1,69 МПа <sup>2</sup>
Довірчий інтервал для математичного сподівання	(3.12)	18,74 ... 20,42 МПа
Імовірна відносна похибка оцінки математичного сподівання	(3.13)	4,26%
Обсяг вибірки, необхідний для оцінювання середнього значення з точністю 5%	(3.14)	15
Довірчий інтервал для дисперсії	(3.15)	2,43 ... 7,97 МПа <sup>2</sup>
Відносна похибка оцінювання дисперсії	(3.16)	53,2%
Мінімально необхідний обсяг вибірки для оцінювання дисперсії з точністю 5%	(3.17)	2152
Довірчий інтервал для стандарту	(3.18)	1,56 ... 2,83 МПа
Відносна похибка оцінювання стандарту	(3.16)	26,6%
Мінімально необхідний обсяг вибірки для оцінювання стандарту з точністю 5%	(3.17)	538

Відносна похибка стандарту прийнята удвічі меншою від похибки оцінювання дисперсії (3.16). Мінімально необхідний обсяг вибірки для оцінювання стандарту з точністю 5% визначений за формулою (3.17) для дисперсії, але з підстановкою допустимого значення точності у 10%. Такий наближений спосіб витікає з наведеної вище вказівки, що відносна похибка дисперсії приблизно удвічі перевищує відносну похибку стандарту.

З отриманих результатів видно, що середнє значення межі міцності бетону, тобто його марка, досить точно визначене за наявною вибіркою з  $N = 20$  даних. Величина стандарту за цією вибіркою визначається з дуже великою імовірною похибкою в 27%, а для забезпечення 5% точності необхідна вибірка з 538 значень межі міцності.

### **3.5. Вибір закону розподілу за результатами статистичного аналізу дослідних даних**

Числові характеристики та гістограма розподілу вибірки дозволяють описати дослідні дані теоретичним законом розподілу. Для цього необхідно вибрати його вид, визначити параметри та перевірити узгодженість із дослідними даними. Вид теоретичного розподілу обирається з урахуванням таких факторів:

- фізичне обґрунтування дослідженого явища (наприклад, нормальний розподіл придатний для опису суми багатьох випадкових величин, логнормальний – добутку випадкових величин, розподіл Гумбеля добре описує максимальні елементи вибірок);
- досвід попередніх досліджень (наприклад, відомо, що розподіли технічних характеристик будівельних матеріалів добре описуються нормальним законом розподілу, розподіли річних чи місячних максимумів кліматичних навантажень відповідають закону Гумбеля, а результати строкових вимірювань швидкості вітру узгоджуються з розподілом Вейбулла);
- вигляд гістограми розподілу, якій повинна відповідати форма кривої густини імовірності обраного теоретичного закону розподілу;
- співвідношення числових характеристик, яке може служити додатковим свідченням придатності обраного закону розподілу (наприклад, коефіцієнт асиметрії розподілу Гумбеля дорівнює  $A=1,14$ , а для експоненціального розподілу характерні значення коефіцієнта варіації  $V = 1$  та коефіцієнта асиметрії  $A = 2$ ).

Для вибору можливого теоретичного розподілу можна скористатися описом часто вживаних законів розподілу, зробленим у підрозділі 2.4. Наведена там інформація (математичне обґрунтування законів розподілу, області визначення ти форми кривих густини розподілів, співвідношення їх числових характеристик, можливості та зручність використання тощо) дозволяє досить обґрунтовано висунути гіпотезу щодо можливого виду теоретичного закону розподілу для імовірнісного опису тієї чи іншої випадкової величини.

Параметри обраного закону розподілу визначаються за методом моментів, який забезпечує рівність числових характеристик теоретичного розподілу та проаналізованої вибірки. Формули та способи визначення параметрів поширених законів розподілу наведені при описі цих законів у підрозділі 2.4. Параметри інших законів розподілу можна визначити, керуючись вказівками літературних джерел, у яких ці закони описані.

За визначеними параметрами та аналітичним виразом густини імовірності обраного розподілу обчислюються значення густини для ряду значень випадкової величини (як правило, для меж інтервалів побудованої гістограми розподілу) і на гістограму наноситься графік густини імовірності. Візуальна відповідність теоретичної кривої густини імовірності до дослідної гістограми є першим орієнтовним свідченням відповідності обраного закону розподілу до дослідних даних та правильності визначення його параметрів.

Остаточне рішення щодо можливості застосування обраного закону розподілу приймається за результатами перевірки його узгодженості з дослідними даними. Для цього використовують спеціальні статистичні критерії узгодженості, найпоширенішим із яких є **критерій Пірсона** хі-квадрат. Згідно з вказівками [1, 2, 4], статистику критерію Пірсона обчислюють за формулою

$$\chi^2 = N \sum_{j=1}^L \frac{(P_j - P_{jt})^2}{P_{jt}}, \quad (3.19)$$

де  $L$  – кількість інтервалів у гістограмі;

$N$  – обсяг вибірки, за якою побудована гістограма розподілу;

$P_j$  – емпірична імовірність попадання випадкової величини до  $j$ -того інтервалу, визначена за гістограмою розподілу;

$P_{jt}$  – теоретична імовірність попадання випадкової величини до  $j$ -того інтервалу, визначена за обраним законом розподілу.

Теоретична імовірність попадання випадкової величини до  $j$ -того інтервалу визначається за формулою

$$P_{jt} = F(C_j) - F(C_{j-1}), \quad (3.20)$$

де  $F(C_j)$  та  $F(C_{j-1})$  – значення функції розподілу обраного теоретичного закону з параметрами, що відповідають проаналізованій вибірці.

Необхідні значення функції нормального та логнормального законів розподілу обчислюється з використанням таблиць з додатку А, книг [1, 2, 4], або функції Excel  $NORM.P.AS.P(x, M, S, I)$  згідно з підрозділом 2.4. Для інших законів розподілу обчислення ведуться за аналітичними виразами інтегральних функцій розподілу, наведеними в підрозділі 2.4.

Обчислення за формулою (3.19) реалізуються в середовищі Excel у вигляді таблиці з такими стовпцями: номери інтервалів  $j$ , межі інтервалів  $C_j$ , емпіричні імовірності попадання випадкової величини до кожного інтервалу  $P_j$ , значення теоретичної функції розподілу для меж інтервалів  $F_{jt}$ , імовірності попадання випадкової величини до кожного інтервалу  $P_{jt}$ , складові формули (3.19), що стоять під знаком суми, для кожного з інтервалів гістограми. Обчислене за формулою (3.19) вибіркове значення статистики  $\chi^2$  порівнюється з критичним значенням  $\chi_{cr}^2(\alpha, k)$ , визначеним за таблицею розподілу Пірсона з [1, 2, 4] чи за таблицею А.2, залежно від рівня значимості  $\alpha$  та кількості ступенів свободи  $k$ .

Рівень значимості  $\alpha$  дорівнює імовірності відкидання правильної гіпотези (визнання невідповідним теоретичного розподілу, який насправді узгоджується з дослідними даними). В статистичних дослідженнях часто приймають рівень значимості  $\alpha=0,05$ . При  $\alpha<0,01$  слід констатувати явну невідповідність обраного закону розподілу до дослідних даних.

Кількість ступенів свободи  $k$  дорівнює кількості інтервалів в гістограмі розподілу  $L$ , зменшеній на кількість параметрів вибраного теоретичного закону розподілу і ще на одиницю. Для нормального, логнормального, закону рівномірного розподілу, а також для розподілів Гумбеля та Вейбулла кількість ступенів свободи дорівнює  $k=L-3$ ; для експоненціального розподілу  $k=L-2$ .

Якщо  $\chi^2 \leq \chi_{cr}^2$ , вибраний закон розподілу **не суперечить** дослідним даним і його можна використовувати для імовірнісного опису дослідженої випадкової величини. При  $\chi^2 > \chi_{cr}^2$  для імовірнісного опису дослідних даних слід вибрати закон розподілу іншого виду.

При масовій перевірці багатьох вибірок однорідних випадкових величин (наприклад, результати дослідження робочих параметрів машини в різний час і в різних умовах експлуатації, результати метеорологічних спостережень на багатьох метеостанціях тощо), які описуються одним і тим же законом розподілу з різними параметрами, процедуру перевірки можна змінити таким чином. За числовими характеристиками кожної з вибірок визначаються параметри закону розподілу і за формулою (3.19) обчислюється статистика критерію Пірсона. З урахуванням відповідної кількості ступенів свободи встановлюється рівень значимості, на якому підібраний теоретичний розподіл не суперечить дослідним даним. Це можна зробити підбором за таблицею розподілу Пірсона [1, 2, 4] або обчислити з використанням функції Excel:  $\alpha = 1 - \text{CHI2.PACП}(\chi^2; k; 1)$ . Після виконання обчислень для усіх вибірок підраховують кількість розподілів, які не суперечать дослідним даним на різних рівнях значимості  $\alpha$ . Якщо підібраний теоретичний закон розподілу на досить високих рівнях значимості не суперечить переважній більшості проаналізованих гістограм і лише в окремих випадках не узгоджується з гістограмами, цей закон розподілу можна вважати загалом придатним для імовірнісного опису дослідженої випадкової величини.

Продовжуючи аналіз *прикладу обробки даних* по межі міцності бетону, відмітимо, що характеристики міцності конструкційних матеріалів зазвичай підпорядковуються нормальному закону розподілу. Можливість його використання для імовірнісного опису даних розглянутого прикладу підтверджується симетричним пагорбоподібним виглядом гістограми, зображеної на рисунку 3.1. В останній колонці таблиці 3.2 наведені значення густини нормального розподілу, обчислені за формулою (2.17). З рисунка 3.1 видно, що побудована за цими значеннями крива густини нормального розподілу добре відображає характер дослідної гістограми. Це дозволяє висунути гіпотезу про можливість використання нормального закону розподілу для імовірнісного опису межі міцності бетону.

Перевірка узгодженості нормального розподілу з наявними дослідними даними виконана за критерієм Пірсона. хід і результати перевірки наведені в таблиці 3.3.

Таблиця 3.3 – Перевірка узгодженості нормального розподілу з дослідною гістограмою розподілу межі міцності бетону

Номер інтервалу $J$	Кінець інтервалу $C_j$	$P_j$	$F_j(t)$	$P_{jt}$	$\chi^2$
	14		0,0071		
1	16	0,050	0,0577	0,0506	0,0001
2	18	0,200	0,2434	0,1857	0,0220
3	20	0,350	0,5729	0,3295	0,0256
4	22	0,250	0,8561	0,2832	0,0778
5	24	0,100	0,9739	0,1179	0,0541
6	26	0,050	0,9976	0,0237	0,5852
С у м и :		1,000		0,9905	0,7648

У результаті обчислень за формулою (3.19) в таблиці 3.3 отримано вибіркоче значення статистики критерію Пірсона  $\chi^2 = 0,765$ . За таблицею А.2 критичних значень розподілу Пірсона для рівня значимості  $\alpha = 0,05$  та кількості ступенів свободи  $k = 6 - 3 = 3$  встановлено критичне значення критерію Пірсона  $\chi_{cr}^2 = 7,82$ . Оскільки  $\chi^2 = 0,765 < \chi_{cr}^2 = 7,82$ , слід вважати, що нормальний розподіл з  $M = 19,58$  МПа та  $S = 2,28$  МПа не суперечить дослідним даним і може використовуватися для імовірнісного опису межі міцності бетону.

Користуючись отриманим нормальним законом розподілу, за формулою (2.19) визначимо клас бетону, як мінімальне значення межі міцності із забезпеченістю 0,95:

$$B = R_0 = M - t \times S = 19,58 - 1,64 \times 2,28 = 15,85 \text{ МПа,}$$

де  $M$  і  $S$  – середнє значення та стандарт за результатами статистичної обробки вибірки даних;

$t = 1,64$  – квантиль нормованого нормального розподілу з таблиці А.2, який відповідає забезпеченості класу бетону  $P = 0,95$ .

Отримані результати показують, що фактична середня міцність бетону  $M = 19,58$  МПа на 2% нижча від необхідної для марки М 200. Не дивлячись на це, гарантована границя міцності бетону  $R_n = 15,85$  МПа відповідає класу В 15, який повинен мати бетон марки М 200. Це можна пояснити тим, що за рахунок добре налагодженої технології виготовлення бетонної суміші фактичний розкид міцності бетону з коефіцієнтом варіації  $V = 11,6\%$  отримано меншим від нормативного значення  $V = 13,5\%$ .

### 3.6. Математичне забезпечення статистичної обробки вибірок випадкових величин в середовищі Microsoft Excel

У попередніх підрозділах описане використання статистичних функцій Microsoft Excel для обробки вибірок випадкових величин. Наведена нижче зведена таблиця містить перелік часто вживаних статистичних функцій Microsoft Excel з їх призначенням та коротким описом параметрів. Інші функції для статистичної обробки вибірок випадкових величин можна знайти в меню вставки функцій Microsoft Excel в розділі "Статистические". При виборі імені функції надається короткий опис її призначення, форми запису та перелік параметрів, що дозволяє обрати необхідну функцію.

Запис функції	Призначення функції та опис параметрів
<i>МИН(список)</i>	Знаходить мінімальний елемент вибірки, заданої в полі комірок "список"
<i>МАКС(список)</i>	Знаходить максимальний елемент вибірки, заданої в полі комірок "список"
<i>СЧЕТ(список)</i>	Підраховує обсяг вибірки, заданої в полі комірок "список"
<i>СРЗНАЧ(список)</i>	Обчислює середнє значення вибірки, заданої в полі комірок "список"
<i>СТАНДОТКЛОН(список), СТАНДОТКЛОН.В(список)</i>	Обчислює стандарт вибірки, заданої в полі комірок "список"
<i>ДИСП.В(список)</i>	Обчислює дисперсію вибірки, заданої в полі комірок "список"
<i>СКОС(список)</i>	Обчислює коефіцієнт асиметрії вибірки, заданої в полі комірок "список"
<i>ЭКЦЕСС(список)</i>	Обчислює коефіцієнт ексцесу вибірки, заданої в полі комірок "список"
<i>МЕДИАНА(список)</i>	Визначає медіану вибірки (значення, яке ділить вибірку на дві рівні частини), заданої в полі комірок "список"
<i>НОРМРАСП(Х, М, S, I)</i>	Обчислює інтегральну функцію (при $I = 1$ ) або густину (при $I = 0$ ) нормального розподілу з математичним сподіванням $M$ та стандартом $S$ для значення аргументу $X$
<i>НОРМОБР(F, М, S)</i>	Обчислює значення аргументу нормального розподілу з математичним сподіванням $M$ та стандартом $S$ , що відповідає заданому значенню інтегральної функції $F$

Запис функції	Призначення функції та опис параметрів
<i>НОРМ.СТ.РАСП(X;I)</i>	Обчислює інтегральну функцію (при $I = 1$ ) або густину (при $I = 0$ ) стандартного нормованого нормального розподілу з математичним сподіванням $M=0$ та стандартом $S=1$ для значення аргументу $X$
<i>НОРМ.СТ.ОБР(F)</i>	Обчислює значення аргументу стандартного нормованого нормального розподілу з математичним сподіванням $M=0$ та стандартом $S=1$ , що відповідає заданому значенню інтегральної функції $F$
<i>ХИ2.РАСП(<math>\chi^2</math>;k;I)</i>	Обчислює інтегральну функцію (при $I = 1$ ) або густину (при $I = 0$ ) розподілу хі-квадрат при $k$ ступенях свободи. Рівень значимості критерію Пірсона $\alpha = 1 - \text{ХИ2.РАСП}(\chi^2; k; I)$
<i>ЧАСТОТА(список, C)</i>	Визначає кількість елементів вибірки, заданої в полі комірок "список", які є меншими від $C$
<i>ГАММА(x)</i>	Обчислює гамма-функцію

## 4. АПРОКСИМАЦІЯ ЕМПІРИЧНИХ ЗАЛЕЖНОСТЕЙ

### 4.1. Кореляція та лінійна регресія випадкових величин

Об'єкти досліджень часто описуються сукупністю декількох ознак, які відображають різні властивості одного й того ж об'єкта. Наприклад, кожен зразок будівельного матеріалу має певну густину, пористість, теплопровідність, межу міцності та інші технічні характеристики. Всі ці показники є випадковими величинами, значення яких змінюються від зразка до зразка, але при цьому можуть бути взаємно залежними. При зростанні пористості матеріалу його середня густина, теплопровідність і межа міцності повинні зменшуватися, що й обумовлює наявність залежностей між цими характеристиками.

Розглянемо дві випадкові змінні  $X$  і  $Y$ , отримані в результаті парного експерименту, тобто виміряні на одному й тому ж об'єкті. Якщо між ними існує *функціональна залежність*, то кожному конкретному значенню  $X$  відповідає конкретне детерміноване (точно визначене) значення  $Y$ . *Стохастична залежність* полягає в тому, що конкретному значенню  $X$  відповідає ряд випадкових значень  $Y$ , які мають певний статистичний розкид. Якщо набір можливих значень (розподіл) випадкової величини  $Y$  змінюється при зміні значення  $X$ , такі випадкові величини називають *залежними*. Так, маса тіла людини не залежить від порядкового номера в алфавітному списку, але залежить від її віку та зросту.

Результати парного експерименту можна зобразити графічно, відклавши вздовж однієї осі значення випадкової величини  $X$ , а вздовж іншої – значення  $Y$ . Такий графік можна побудувати в середовищі Microsoft Excel як діаграму типу "точечная". Кожна пара даних, яка характеризує один і той же об'єкт (зріст і маса тіла певної людини, середня густина та межа міцності певного зразка легкого бетону тощо), утворює на графіку одну точку. За таким графіком можна візуально встановити наявність і характер залежності між проаналізованими випадковими величинами.

Статистичною мірою тісноти лінійного (або близького до лінійного) зв'язку між двома випадковими величинами є *коефіцієнт кореляції*, який обчислюється за формулою

$$R_{xy} = \frac{1}{(N-1) \times S_x \times S_y} \sum_{i=1}^N (X_i - M_x)(Y_i - M_y), \quad (4.1)$$

де  $N$  – обсяг кожної з вибірок;

$X_i$  і  $Y_i$  – відповідні елементи вибірок, які відображають результати парного експерименту для одного і того ж об'єкта;

$M_X$  і  $M_Y$  – оцінки математичних сподівань (середні значення) обох вибірок;

$S_X$  і  $S_Y$  – оцінки стандартів обох вибірок.

При статистичній обробці даних в середовищі електронних таблиць Microsoft Excel коефіцієнт кореляції зручно обчислювати з використанням функції *KORPEЛ*(масив\_X; масив\_Y), у якій масив\_X; масив\_Y – виділені стовпці чи рядки, що містять значення змінних X та Y.

За значенням коефіцієнта кореляції можна судити про наявність і характер зв'язку між випадковими величинами. Коефіцієнт кореляції незалежних випадкових величин, зображених на рисунку 4.1 а, наближається до нуля. Відмінність коефіцієнта кореляції від нуля свідчить про наявність зв'язку між випадковими величинами, а його близькість за модулем до одиниці – про наближеність такого зв'язку до лінійного функціонального. При  $R_{XY} > 0$  (рисунок 4.1 б) збільшення величини X викликає зростання Y, а при  $R_{XY} < 0$  (рисунок 4.1 в), навпаки, більшим значенням X відповідають менші Y.

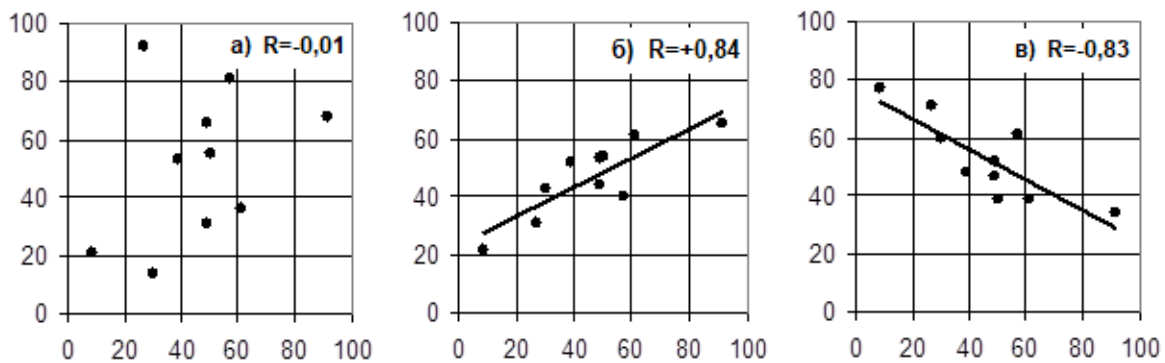


Рис. 4.1. Приклади залежностей між випадковими величинами

Якщо між проаналізованими величинами існує досить тісний зв'язок (вважається, що при значеннях коефіцієнта кореляції  $|R_{XY}| > 0,7$ ), залежність між ними доцільно описувати *рівнянням лінійної регресії* виду

$$Y = A + B \times X \quad (4.2)$$

Параметри  $A$ ,  $B$  апроксимуючої лінійної функції (4.2) визначаються *методом найменших квадратів*, який забезпечує мінімізацію суми квадратів відхилень дослідних значень  $Y$  від апроксимуючої прямої (4.2). Для цього вони обчислюються за формулами

$$A = M_Y - B \times M_X; \quad B = R_{XY} \times S_Y / S_X, \quad (4.3)$$

усі позначення в яких відповідають (4.1). При використанні електронних таблиць Microsoft Excel параметри  $A$ ,  $B$  лінійного рівняння регресії (4.2) можна обчислити за формулами (4.3) або за вбудованими функціями  $ОТРЕЗОК(масив\_Y; масив\_X)$  та  $НАКЛОН(масив\_Y; масив\_X)$ . Параметри  $A$ ,  $B$  можна також побачити на побудованій діаграмі типу "точечная", виконавши команду "вставить линию тренда" з встановленою міткою "показать уравнение на диаграмме".

Із рисунків 4.1 б і 4.1 в видно, що апроксимуючі прямі (4.2) описують дослідні залежності „у середньому”, а реальні дослідні значення  $Y$  випадковим чином відхиляються від апроксимуючої прямої. Зростання модуля коефіцієнта кореляції до одиниці зменшує величину розкиду до нуля, тобто перетворює стохастичну залежність у функціональну.

Рівняння регресії виду (4.2) використовують для прогнозування невідомих значень величини  $Y$  через відомі значення  $X$ . Це рівняння дає середню оцінку, від якої реальні значення  $Y$  можуть істотно відхилитися в більший та в менший бік. На практиці можуть виникнути ситуації, коли необхідно прогнозувати можливі значення функції  $Y$  в запас надійності, тобто в бік завищення чи заниження залежно від їх фізичного смислу. Інакше кажучи, зображені на рисунках 4.1 б і 4.1 в лінії регресії слід провести поблизу верхньої чи нижньої межі дослідних точок таким чином, щоб переважна більшість дослідних точок (наприклад, 80%, 90% чи 95%) знаходилася в межах безпечної області. Цю проблему можна розв'язати шляхом підбору параметра  $A$  в рівнянні (4.2), зміни якого паралельно зміщують лінію зв'язку угору чи вниз.

В середовищі Microsoft Excel поставлену задачу можна розв'язати таким графічним способом:

- на збудований графік залежності типу рисунка 4.1 додати ще один ряд даних, у якому відображаються значення, обчислені за рівнянням (4.2) з параметрами  $A$ ,  $B$ , заданими в окремих комірках;
- змінюючи параметр  $A$  (можливо також  $B$ ), переміщувати графік нової лінійної залежності таким чином, щоб добитися необхідного положення поблизу верхньої чи нижньої межі дослідних точок;
- необхідне положення нової прямої обирається на око, або шляхом підрахунку кількості дослідних точок, для яких прогноз значень  $Y$  через  $X$  виконується не в запас надійності.

Більш детальне роз'яснення викладеного способу наведено в прикладах з розділу 6.

## 4.2. Опис нелінійних залежностей типовими функціями

Емпіричні залежності між властивостями об'єктів дослідження, отримані шляхом парних експериментів, часто мають нелінійний характер. Такі залежності описують (апроксимують) різноманітними аналітичними функціями, параметри яких визначають *методом найменших квадратів*. В основу методу найменших квадратів покладено наступний загальний принцип: параметри апроксимуючої функції підбираються таким чином, щоб сума квадратів відхилень дослідних точок (значень функції  $Y$ ) від апроксимуючої (теоретичної) кривої була мінімально можливою.

При *виборі виду апроксимуючої функції* слід ураховувати вигляд графіка залежності, точність апроксимації, асимптотичну поведінку апроксимуючої функції та зручність її подальшого використання.

*Точність апроксимації* оцінюється ступенем розкиду дослідних точок відносно підбраної функції. Для цього потрібно обчислити абсолютні чи відносні відхилення дослідних точок від апроксимуючої функції та виконати статистичну обробку цих відхилень, як вибірки випадкової величини. При правильному підборі параметрів теоретичної функції середнє значення цих відхилень завжди буде близьким до нуля, а стандарт може служити мірою точності апроксимації. Задавши певний рівень забезпеченості, за формулою (2.19) можна визначити імовірні межі абсолютних чи відносних відхилень дослідних точок від апроксимуючої кривої. Так наприклад, при  $t = 1,64$  формула (2.19) дає двосторонні довірчі межі з забезпеченістю 0,9. У цьому випадку похибка апроксимації не перевищує встановлених за формулою (2.19) меж у 90% випадків.

*Асимптотична поведінка* апроксимуючої функції при зменшенні чи зростанні аргументу повинна відповідати фізичній природі дослідженого явища. Наприклад, зменшення перерізу деталі чи елемента конструкції до нуля повинно викликати нескінчене зростання напружень; при зростанні кількості циклів навантаження допустимі напруження в сталі повинні асимптотично наближатися до межі витривалості.

*Зручність використання* апроксимуючої функції визначається її подальшим призначенням. Якщо функція буде використовуватися для ручних інженерних розрахунків, вона повинна мати достатньо простий вигляд та округлені значення параметрів, а також по можливості не містити складних математичних функцій. Якщо апроксимуюча функція буде використовуватися в подальших дослідженнях, вона має забезпечити можливість виконання математичних операцій, які з нею доведеться здійснювати (алгебраїчні дії, диференціювання, інтегрування тощо).

У практиці апроксимації нелінійних емпіричних залежностей часто використовується сімейство двопараметричних функцій, які мають досить прості аналітичні вирази, цілком визначену асимптотичну поведінку та можуть описувати нелінійні залежності різноманітного характеру:

$$\begin{aligned} Y &= A + B \ln(X); & Y &= A e^{B X}; & Y &= A X^B; \\ Y &= X/(A + B X); & Y &= A + B X^K; & Y &= A + B/X. \end{aligned} \quad (4.4)$$

Параметри перших трьох функцій (4.4) можна підібрати в середовищі Microsoft Excel. Для цього необхідно на побудованій діаграмі типу "точечная" виконати команду "добавить линию тренда", встановити мітку "показывать уравнение на диаграмме" і вибрати відповідний вид функції.

Параметри трьох останніх функцій (4.4) обчислюються за методом найменших квадратів шляхом чисельної мінімізації суми квадратів відхилень дослідних точок (дослідних значень функції  $Y$ ) від апроксимуючої функції. При використанні табличного процесора Microsoft Excel для цього необхідно виконати такі операції:

- занести на робочий аркуш вихідні дані (значення  $X_D$  та  $Y_D$  по кожній експериментальній точці) у вигляді двох стовпців або рядків;
- у третьому стовпчику (чи рядку) для значень аргументу  $X_D$ , для яких наявні дослідні значення функції  $Y_D$ , записати формули для обчислення значень апроксимуючої функції  $Y_T$  при початкових параметрах  $A$ ,  $B$ , заданих в окремих комірках електронної таблиці;
- за функцією  $\text{СУММКВРАЗН}(\text{масив\_}Y_D, \text{масив\_}Y_T)$  в комірці цільової функції обчислити суму квадратів відхилень дослідних значень функції  $Y_D$  від апроксимуючих значень  $Y_T$ ;
- скористатися функцією Microsoft Excel "поиск решения", яка здійснює підбір значень параметрів  $A$ ,  $B$ , при яких сума квадратів відхилень в комірці цільової функції буде мінімально можливою.

Функція Excel "поиск решения" реалізує розв'язання такої задачі: змінюючи значення комірок з величинами параметрів  $A$ ,  $B$ , зробити значення цільової функції (суму квадратів відхилень дослідних значень  $Y_D$  від значень апроксимуючої функції  $Y_T$  з параметрами  $A$ ,  $B$  для усіх дослідних точок) мінімально можливою. У результаті успішного пошуку виділені комірки міститимуть підібрані значення параметрів  $A$ ,  $B$  апроксимуючої функції. У випадку невдалого пошуку слід повторити його при інших початкових значеннях параметрів  $A$ ,  $B$ .

У практиці аналізу даних зустрічаються залежності, у яких при різних значеннях аргументу  $X_D$  значення функції  $Y_D$  можуть відрізнятися на порядки, тобто в десятки чи сотні разів. Малим значенням функції зазвичай відповідають пропорційні їм малі значення відхилень дослідних даних від аналітичної залежності, тому ці відхилення практично не впливають на загальну суму квадратів відхилень. Підібрана апроксимуюча функція буде добре описувати частину залежності з великими значеннями дослідної функції і даватиме великі відносні похибки апроксимації в зоні дослідних точок з малими значеннями  $Y_D$ . У такому випадку для забезпечення рівномірної по усіх значеннях аргументу  $X_D$  відносної точності апроксимації краще мінімізувати не суму квадратів відхилень дослідних значень функції  $Y_D$  від апроксимуючих значень  $Y_T$ , а суму квадратів відсотків цих відхилень. Такий підхід вимагає змінити викладений вище алгоритм визначення параметрів апроксимуючої функції наступним чином:

- в розрахунковій таблиці формується четвертий стовпчик, в комірках якого обчислюються відсотки відхилень  $\Delta$  дослідних значень функції  $Y_D$  від апроксимуючих значень  $Y_T$  при поточних значеннях параметрів  $A, B$  апроксимуючої функції;
- в комірці цільової функції обчислюється сума квадратів відсотків відхилень дослідних значень функції  $Y_D$  від апроксимуючих значень  $Y_T$  з використанням функції *СУММКВ (масив\_Δ)*.

У результаті мінімізації обчисленої таким чином цільової функції з використанням функції Microsoft Excel "поиск решения" будуть отримані значення параметрів  $A, B$ , які забезпечать рівномірність відносних похибок апроксимації в усьому діапазоні значень аргументу дослідженої залежності. Відносні похибки апроксимації для кожної з дослідних точок містяться в четвертому стовпчику розрахункової таблиці.

Microsoft Excel передбачає також можливість апроксимації різних залежностей *алгебраїчним поліном* виду

$$Y = A_0 + A_1 X + A_2 X^2 + \dots + A_K X^K . \quad (4.5)$$

Алгебраїчний поліном (4.5) є досить універсальною і зручною для використання функцією, яка часто вживається для опису різноманітних емпіричних залежностей. За ним легко виконувати обчислення, його можна диференціювати, інтегрувати та виконувати інші перетворення. Значна кількість екстремумів (на одиницю менша від ступеня полінома  $K$ ) зумовлює гнучкість отриманої кривої; тому на практиці рідко доводиться

використовувати поліноми вище від 5...6 ступеня. Підбір параметрів  $A_0...A_K$  полінома (4.5) здійснюється командою "добавить линию тренда" в діаграмі типу "точечная". Основним недоліком алгебраїчного полінома (4.5) є непередбачуваність його асимптотичної поведінки, що не дозволяє використовувати його поза межами області зміни аргументу дослідних точок, використаних для апроксимації.

### 4.3. Опис нелінійних та багатофакторних залежностей довільними функціями

Для апроксимації нелінійних залежностей між властивостями об'єктів дослідження окрім функцій, описаних у попередньому підрозділі, можуть використовуватися довільні функції, вид яких обирається за виглядом експериментально отриманих графіків з урахуванням фізичних уявлень та зручності користування цими функціями. Це можуть бути комбінації різних елементарних функцій, кожна з яких вносить свій вклад у поведінку результуючої залежності. Наприклад, відомий вираз для опису періодично затухаючої автокореляційної функції випадкового процесу має вигляд:  $R(\tau) = \exp(-\alpha \cdot \tau) \cdot \cos(\omega \cdot \tau)$ . Косинус забезпечує періодичність цієї функції, а експонента – поступове затухання з наближенням до нуля.

Підбір параметрів такого роду складних функцій теж можна виконати методом найменших квадратів у середовищі Microsoft Excel, як це описано у попередньому підрозділі для двопараметричних функцій (4.4). Залежно від характеру дослідних даних, у якості цільової функції можна використовувати суму квадратів відхилень дослідних значень функції  $Y_d$  від апроксимуючих значень  $Y_T$ , або суму квадратів відсотків цих відхилень. Основна складність полягає у вдалому виборі аналітичного виразу апроксимуючої функції шляхом теоретичного чи логічного обґрунтування або перебору декількох можливих варіантів.

Поряд з парними експериментами, можуть виконуватися більш складні дослідження з метою виявити й описати залежність результуючої величини не від одного, а одночасно від декількох впливаючих на неї аргументів. Результатом таких досліджень є табличні залежності однієї змінної від  $K$  аргументів:  $Y = f(X_1, X_2, \dots, X_K)$ . У випадку двох аргументів отримуємо таблицю з двома входами, яку можна наочно представити у вигляді поверхні. Графіки такого виду для випадку систематичної зміни аргументів  $X_1$  та  $X_2$  з певними кроками можна побудувати графічними

засобами Microsoft Excel у формі поверхні чи ліній рівня, що відображають цю поверхню подібно горизонталям на географічній карті. При більшій кількості аргументів результати експерименту слід подавати у формі таблиці, кожен рядок якої містить комплекс властивостей одного об'єкту дослідження: значення усіх аргументів  $X_1, X_2 \dots X_K$  і результуючої функції  $Y$ . Кількість рядків цієї таблиці дорівнює кількості обстежених об'єктів, тобто кількості експериментальних точок у багатовимірному просторі.

Параметри багатофакторних апроксимуючих функцій обчислюються в середовищі Microsoft Excel таким же способом, як і параметри функції одного аргументу. Найбільшу складність викликає вибір виду функції декількох аргументів, що можна зробити таким чином:

- проаналізувати фізичну сутність дослідженого явища та встановити якого типу залежностями воно повинно описуватися;
- збудувати і проаналізувати парні залежності результуючої величини  $Y$  окремо від кожного з аргументів  $X_1, X_2 \dots X_K$ , при фіксованих значеннях інших аргументів, виявити форму цих залежностей та можливий вид апроксимуючих функцій;
- спробувати перемножити або скласти функції залежностей від окремих аргументів та підібрати параметри отриманої об'єднаної функції.

Для не надто складних залежностей досить ефективним може бути використання алгебраїчних поліномів, які містять аргументи в різних ступенях та їх добутки. Наприклад, для випадку двох аргументів  $X_1$  та  $X_2$  можна записати квадратний поліном такого виду:

$$Y = A_0 + A_1 X_1 + A_2 X_1^2 + A_3 X_2 + A_4 X_2^2 + A_5 X_1 X_2 . \quad (4.6)$$

Можна спробувати підібрати кубічний поліном аналогічного типу, який матиме вже не 5, а 9 коефіцієнтів, що підлягають визначенню за дослідними даними. У випадку трьох аргументів перебір усіх можливих комбінацій аргументів до другого чи третього ступеню дає набагато складнішу функцію. Слід урахувати, що кількість дослідних точок повинна бути не меншою за кількість параметрів апроксимуючої функції, які підбираються за дослідними даними.

Після успішного підбору параметрів за допомогою функції Excel "поиск решения" необхідно спробувати округлити отримані значення параметрів до прийнятних величин (2...4 значимі цифри) та взагалі відкинути параметри, які мало впливають на результуючу функцію. Вилучення частини параметрів спрощує вид підбраної апроксимуючої функції та користування нею.

Округлення значень та вилучення параметрів здійснюють по одному, контролюючи при цьому зміни суми квадратів відхилень дослідних точок від підбраної залежності, яка є цільовою функцією при пошуку рішення. Бажано також відслідковувати зміни максимальної похибки апроксимації. Для вилучення параметра не треба змінювати формулу для обчислення функції, уведена в комірки Excel, достатньо встановити у відповідній комірці значення параметра рівним нулю. Якщо цільова функція змінюється неістотно, зроблені зміни можна прийняти. Зауважимо, що величина параметра (коефіцієнта при незалежній змінній) залежить від величини змінної, яка на нього множиться. Тому коефіцієнт 0,01 при  $X^3$  може бути більш значимим, ніж коефіцієнт 1,0 при  $X$ .

Більш детально методика апроксимації дослідних даних довільними функціями одного та декількох аргументів пояснена на реальних прикладах у розділі 6.

#### **4.4. Узгодженість теоретичних результатів з дослідними даними**

Необхідність у порівнянні теорії з експериментальними даними виникає тоді, коли теорія розроблена незалежно від експериментальних даних. Перевірка адекватності теорії здійснюється шляхом порівняння теоретичних та експериментальних значень результуючого (контрольного) параметра, з метою визначення якого розроблена теорія. Наприклад, для підтвердження нової методики розрахунку слід порівняти зусилля чи напруження в деталях машин чи в елементах будівельних конструкцій, які отримані шляхом розрахунку за цією методикою, з тими ж значеннями, отриманими в результаті випробувань. По суті це порівняння є парним експериментом, результати якого аналізують статистичними методами.

Експериментальні дані завжди мають певний розкид, обумовлений як похибками вимірювань, так і природною мінливістю вимірюваних параметрів об'єктів дослідження (розмірів, технічних характеристик матеріалів тощо). Тому результати теоретичних розрахунків принципово не можуть точно співпадати з експериментальними даними. Залежно від фізичного смислу задачі та мети дослідження, теоретичні значення результуючого (контрольного) параметра повинні або наближатися до експериментальних даних в середньому, або розміщуватися на верхній чи нижній межі дослідних точок, створюючи таким чином необхідний запас надійності при використанні розробленої теорії.

Об'єктом аналізу відповідності теорії та експерименту є абсолютні чи відносні відхилення теоретичних результатів від експериментальних значень контрольного параметра. Сукупності цих відхилень для усіх дослідних точок утворюють вибірки випадкових величин абсолютних  $\Delta_{Ai}$  та відносних відхилень  $\Delta_{Bi}$ , елементи яких обчислюється за формулами:

$$\Delta_{Ai} = X_{Ti} - X_{Ei}; \quad \Delta_{Bi} = \frac{X_{Ti} - X_{Ei}}{X_{Ei}} \times 100\%, \quad (4.7)$$

де  $X_{Ti}$  та  $X_{Ei}$  – теоретичне та експериментальне значення контрольного параметра для одного й того ж об'єкта чи набору вихідних даних.

Досвід показує, що в переважній більшості випадків розподіли абсолютних та відносних відхилень (4.7) близькі до нормального закону розподілу, який і використовується в подальшому аналізі. При цьому рекомендується використовувати таку процедуру:

1. Збудувати графік залежності теоретичних значень контрольного параметра від дослідних, візуально проаналізувати розкид даних та обрати вид відхилень для аналізу. Якщо відображені на графіку абсолютні відхилення мало залежать від величини контрольного параметра, вони обираються для аналізу. Якщо з ростом контрольного параметра абсолютні відхилення помітно збільшуються, слід аналізувати відносні відхилення.
2. За однією з формул (4.7) обчислити відхилення, обрані для аналізу (абсолютні чи відносні), для усіх наявних об'єктів дослідження та виконати статистичну обробку отриманої вибірки відхилень теоретичних значень контрольного параметра від результатів експерименту з визначенням середнього значення  $M_{\Delta}$  і  $S_{\Delta}$ .
3. Визначити довірчі межі відхилень згідно з формулою (2.19) для нормального закону розподілу, яка в наведених вище позначеннях набирає такого вигляду:

$$\Delta_{\min} = M_{\Delta} - t \times S_{\Delta}; \quad \Delta_{\max} = M_{\Delta} + t \times S_{\Delta}. \quad (4.8)$$

Величина параметра  $t$  обирається за таблицями нормального розподілу залежно від заданої довірчої імовірності  $P$ . Наприклад, при  $P = 0,9$  слід прийняти  $t = 1,64$ , а для імовірності  $P = 0,95$  значення  $t = 1,96$ .

4. Отримане в результаті статистичної обробки середнє значення  $M_{\Delta}$  можна вважати систематичною похибкою теорії (абсолютною чи відносною залежно від прийнятого критерію порівняння), а довірчі межі  $\Delta_{\min}$  і  $\Delta_{\max}$  – імовірними межами випадкових відхилень теоретичних значень від експериментальних. Іншими словами можна сказати, що

перевірена теорія дає середню похибку  $M_d$  (у середньому завищує чи занижує значення контрольного параметра на  $M_d$  відсотків чи фізичних одиниць вимірювання) а для  $100 \times P\%$  випадків випадкова похибка контрольного параметра не виходить за межі інтервалу  $[\Delta_{min}; \Delta_{max}]$ , заданого у відсотках чи у фізичних одиницях вимірювання.

Існує також практика аналізу відношення теоретичних значень контрольного параметра до експериментальних

$$\Delta_i = X_{Ti} / X_{Ei} . \quad (4.9)$$

Порівняння з формулою (4.7) показує, що відношення (4.9) по суті є відносним відхиленням контрольного параметра. Якщо від (4.9) відняти одиницю і помножити на 100%, отримуємо другу формулу (4.7). Отже, твердження, що відхилення (4.9) має середнє значення 0,94 і з імовірністю 0,9 знаходиться в межах від 0,84 до 1,04 еквівалентне твердженню, що теорія систематично занижує контрольний параметр на 6%, а випадкові відхилення теорії від експерименту для 90% випадків не виходять за межі від -16% до +4%.

Викладена методика дозволяє на базі досить простої процедури отримати коректні висновки щодо адекватності теорії, чітко зрозумілі з інженерної точки зору. Таким же чином можна порівняти між собою результати розрахунків за двома різними теоріями чи результати експериментальних досліджень двох авторів, якщо вони виконувалися в еквівалентних умовах і з фізичної точки зору повинні бути однаковими.

Більш детальні пояснення щодо практичних способів перевірки узгодженості теоретичних результатів з дослідними даними надані в розділі 6 на конкретних прикладах такої перевірки.

#### **4.5. Математичне забезпечення апроксимації емпіричних залежностей в середовищі Microsoft Excel**

У даному розділі надані рекомендації з використання статистичних функцій Microsoft Excel для аналізу зв'язку між випадковими величинами та апроксимації емпіричних залежностей між ними. Наведена нижче зведена таблиця містить перелік цих функцій, їх призначення та короткий опис параметрів. Вказані в таблиці функції для статистичної обробки вибірок випадкових величин призначені для аналізу відхилень дослідних даних від апроксимуючих функцій згідно з вказівками, наведеними в розділі 3. В усіх функціях аргумент "масив" означає сукупність комірок у вигляді стовпчика чи рядка, який містить необхідні аргументи чи функції.

Додаткові функції, які можуть використовуватися для аналізу та опису залежностей між випадковими величинами, можна знайти в меню вставки функцій Microsoft Excel в розділах "Статистические" і "Математические".

Запис функції	Призначення функції та опис параметрів
<i>КОРРЕЛ</i> (масив_X; масив_Y)	Обчислює коефіцієнт кореляції двох вибірок випадкових величин, заданих полями комірок "масив_X" і "масив_Y"
<i>ОТРЕЗОК</i> (масив_Y; масив_X)	Обчислює параметр <i>A</i> рівняння лінійної регресії (4.2)
<i>НАКЛОН</i> (масив_Y; масив_X)	Обчислює параметр <i>B</i> рівняння лінійної регресії (4.2)
<i>СУММКВ</i> (масив_X)	Обчислює суму квадратів чисел, заданих у полі комірок "масив_X"
<i>СУММКВРАЗН</i> (масив_X), масив_Y)	Обчислює суму квадратів попарних різниць елементів двох вибірок випадкових величин, заданих полями комірок "масив_X" і "масив_Y"
<i>СЧЕТ</i> (список)	Підраховує обсяг вибірки, заданої в полі комірок "список"
<i>МИН</i> (список)	Знаходить мінімальний елемент вибірки, заданої в полі комірок "список"
<i>МАКС</i> (список)	Знаходить максимальний елемент вибірки, заданої в полі комірок "список"
<i>СРЗНАЧ</i> (список)	Обчислює середнє значення вибірки, заданої в полі комірок "список"
<i>СТАНДОТКЛОН</i> (список), <i>СТАНДОТКЛОН.В</i> (список)	Обчислює стандарт вибірки, заданої в полі комірок "список"

## 5. БАГАТОВИМІРНІ СТАТИСТИЧНІ МЕТОДИ

### 5.1. Порівняння випадкових величин та дисперсійний аналіз

У практиці досліджень може виникнути необхідність порівняння випадкових величин з метою встановлення їх близькості чи значимих відмінностей. Наприклад, порівнюються результати метеорологічних спостережень на двох сусідніх метеостанціях з метою підтвердження їх близькості та можливості об'єднання в одну більш інформативні вибірку. Порівняння значень межі міцності чи інших характеристик, отриманих за результатами випробувань зразків бетону, виготовлених з використанням певної добавки та без неї, дозволить виявити значиму різницю і таким чином підтвердити вплив суттєвий добавки на ці характеристики.

Рішення про рівність статистичних характеристик випадкових величин приймаються на певному рівні значимості  $\alpha$ , який дорівнює імовірності помилкового відкидання правильної гіпотези (імовірність того, що рівні математичні сподівання чи дисперсії випадкових величин будуть помилково визнані значимо різними). В технічних дослідженнях зазвичай приймають рівень значимості  $\alpha = 0,05$ . У деяких випадках допускають прийняття рішення на рівні значимості  $\alpha = 0,02$  чи навіть  $\alpha = 0,01$ .

Розглянемо найпростіші методи порівняння математичних сподівань та дисперсій двох вибірок, рекомендовані в посібнику [4]. Строго кажучи, описані нижче критерії порівняння математичних сподівань та дисперсій розроблені для аналізу випадкових величин з нормальними законами розподілу. Використання цих критеріїв для пагорбоподібних розподілів, які не сильно відрізняються від нормального, дає наближені результати, але цілком прийнятні для практичних інженерних досліджень.

*Рівність математичних сподівань* двох випадкових величин можна перевірити за таким алгоритмом:

1. За результатами статистичної обробки двох вибірок обсягами  $N_1$ ,  $N_2$  визначаються математичні сподівання  $M_1$ ,  $M_2$  та дисперсії  $D_1$ ,  $D_2$  обох вибірок. Нагадаємо, що дисперсія дорівнює квадрату стандарту

$$D_1 = S_1^2, \quad D_2 = S_2^2. \quad (5.1)$$

2. Статистика критерію обчислюється за формулою

$$Z = |M_1 - M_2| / \sqrt{\frac{D_1}{N_1} + \frac{D_2}{N_2}}, \quad (5.2)$$

3. Обчислена статистика (5.2) порівнюється з двостороннім критичним значенням нормального розподілу  $Z_{cr}$ , яке визначається за таблицями з книг [1, 2, 4] або з додатка А для заданого рівня значимості  $\alpha$ . В середовищі Microsoft Excel критичне значення цього критерію можна обчислити за функцією  $HOPM.CT.OBP(1 - \alpha / 2)$ . Критичні значення  $Z_{cr}$  для часто вживаних рівнів значимості дорівнюють:

Рівень значимості $\alpha$	0,2	0,1	0,05	0,02	0,01
Критичне значення $Z_{cr}$	1,28	1,64	1,96	2,33	2,58

4. При  $Z < Z_{cr}$  можна вважати математичні сподівання проаналізованих випадкових величин практично рівними, тобто  $M_1 \approx M_2$ . При  $Z > Z_{cr}$  математичні сподівання проаналізованих випадкових величин слід вважати значимо різними.

**Рівність дисперсій** двох випадкових величин можна перевірити на основі критерію Фішера за таким алгоритмом:

1. За результатами статистичної обробки двох вибірок обсягами  $N_1$ ,  $N_2$  визначаються їх математичні сподівання  $M_1$ ,  $M_2$  та дисперсії  $D_1$ ,  $D_2$ . Першим номером слід позначити вибірку з більшою дисперсією, оскільки подальші формули записані з урахуванням умови  $D_1 > D_2$ .
2. Статистика критерію обчислюється за формулою

$$F = D_1 / D_2 . \quad (5.3)$$

3. Обчислена статистика (5.3) порівнюється з правостороннім критичним значенням розподілу Фішера  $F_{cr}(\alpha, k_1, k_2)$  для заданого рівня значимості  $\alpha$  та ступенів свободи  $k_1 = N_1 - 1$ ,  $k_2 = N_2 - 1$ . Критичне значення розподілу Фішера можна встановити за таблицями з додатку А чи з книг [2, 4], а в середовищі Microsoft Excel – обчислити функцією  $F.OBP.PX(\alpha; k_1; k_2)$ .
4. При  $F < F_{cr}$  дисперсії обох вибірок можна вважати рівними чи досить близькими, тобто прийняти  $D_1 \approx D_2$ . У випадку  $F > F_{cr}$  дисперсії проаналізованих випадкових величин слід вважати значимо різними.

**Дисперсійний аналіз** використовують при необхідності порівняти математичні сподівання декількох випадкових величин. Така ситуація може виникнути при бажанні об'єднати декілька вибірок даних, отриманих у близьких умовах, або з метою виявлення впливу певного фактора на значення випадкової величини, що досліджується. В якості прикладів можна вказати на задачу об'єднання результатів спостережень на декількох територіально близьких метеостанціях, або дослідження впливу різних технологій виготовлення на технічні характеристики виробів.

Задачу дисперсійного аналізу стосовно порівняння середніх значень декількох вибірок можна сформулювати таким чином. Розглядаються  $L$  вибірок з номерами  $k = 1 \dots L$  та обсягами  $N_k$  результатів спостережень  $X_{ki}$ ,  $i = 1 \dots N_k$ . Індексом  $k$  позначено номер вибірки, а індексом  $i$  – порядковий номер елемента у вибірці. Необхідно виявити значимі відмінності між математичними сподіваннями цих вибірок, або навпаки – підтвердити можливість об'єднання усіх наявних даних в одну спільну вибірку обсягом

$$N = \sum_{k=1}^L N_k . \quad (5.4)$$

Порівняльний аналіз вибірок виконується в такому порядку:

1. Обчислюються середні значення та дисперсії для кожної з вибірок

$$M_k = \frac{1}{N_k} \sum_{i=1}^{N_k} X_{ki} ; \quad D_k = \frac{1}{N_k - 1} \sum_{i=1}^{N_k} (X_{ki} - M_k)^2 . \quad (5.5)$$

2. Обчислюються середнє значення об'єднаної вибірки

$$M = \frac{1}{N} \sum_{k=1}^L \sum_{i=1}^{N_k} X_{ki} ; \quad (5.6)$$

3. Залишкова дисперсія, яка в середньому характеризує випадковий розкид даних в межах однієї вибірки, дорівнює

$$D_3 = \sum_{k=1}^L D_k (N_k - 1) / (N - L) . \quad (5.7)$$

4. Факторна дисперсія, яка відображає розкид середніх значень вибірок і характеризує вплив визначального фактора (номера вибірки) на сумарну дисперсію об'єднаної вибірки, обчислюється за формулою:

$$D_\phi = \sum_{k=1}^L N_k (M_k - M)^2 / (L - 1) . \quad (5.8)$$

5. Статистика критерію Фішера для порівняння дисперсій дорівнює

$$F = D_\phi / D_3 . \quad (5.9)$$

6. Критичне значення статистики (5.9)  $F_{cr}(\alpha, k_\phi, k_3)$  визначається за таблицями F-розподілу для обраного рівня значимості (зазвичай  $\alpha=0,05$  або  $\alpha=0,01$ ) та кількості ступенів свободи  $k_\phi=L-1$  та  $k_3=N-L$ . Як і в попередньому випадку, значення  $F_{cr}$  можна встановити за таблицями додатку А чи з книг [2, 4], а також за статистичною функцією Microsoft Excel  $F.OBP.PX(\alpha; k_\phi; k_3)$ .

7. Якщо отримано  $F \leq F_{cr}$ , проаналізовані вибірки можна вважати однорідними та об'єднати їх в спільну вибірку обсягу (5.4).

8. Значення  $F > F_{cr}$  вказує на значимі відмінності математичних сподівань проаналізованих вибірок та на істотний вплив визначального фактора на досліджені випадкові величини.

Дисперсійний аналіз можна виконати в середовищі Microsoft Excel, скориставшись для цього функцією "Однофакторний дисперсійний аналіз" з меню "Аналіз даних". Для цього необхідно занести усі елементи вихідних вибірок у суміжні стовпці електронної таблиці та увійти в функцію "Однофакторний дисперсійний аналіз". Після виділення області комірок, які містять вихідні вибірки, та встановлення необхідних даних в меню функції усі описані вище обчислення виконуються в автоматичному режимі. В таблиці результатів залишкова дисперсія (5.7) позначена як "источник вариации внутри групп", а факторна дисперсія (5.8) – "источник вариации между группами". Кінцевими результатами розрахунку є значення статистики критерію Фішера (5.9) та критичне значення, за якими і слід робити висновок про близькість проаналізованих вибірок чи значимі відмінності між ними. Детальні пояснення щодо методики проведення дисперсійного аналізу в середовищі Microsoft Excel надані в розділі 6 на прикладі аналізу даних в ваги снігового покриву.

## 5.2. Багатофакторні експерименти та їх аналіз

Природні, технічні чи соціальні об'єкти можуть характеризуватися сукупністю декількох числових ознак, які визначають різні властивості цих об'єктів. Нижче розглядається вибірка з  $N$  об'єктів (спостережень), кожен з яких характеризується сукупністю одних і тих же  $M$  числових ознак. Об'єктами можуть бути окремі вироби, конструкції, географічні чи природні явища, люди тощо. Числові ознаки обираються відповідно до об'єктів і залежно від мети дослідження. Така інформація утворює матрицю

$$\begin{pmatrix} x_{1,1} & \dots & x_{1,j} & \dots & x_{1,M} \\ x_{2,1} & \dots & x_{2,j} & \dots & x_{2,M} \\ x_{i,1} & \dots & x_{i,j} & \dots & x_{i,M} \\ \dots & \dots & \dots & \dots & \dots \\ x_{N,1} & \dots & x_{N,j} & \dots & x_{N,M} \end{pmatrix}, \quad (5.10)$$

Рядки матриці (5.10) відповідають різним об'єктам з номерами  $i = 1, \dots, N$ , а стовпці – різним ознакам з номерами  $j = 1, \dots, M$ . Одна з ознак (стовпців матриці) може вважатися залежною від інших.

Найпростішим варіантом аналізу таких даних є статистична обробка кожного стовпця матриці (5.10) як вибірки випадкової величини, аналіз та опис парних залежностей між ними. Більш узагальнені та інформативні результати можна отримати при застосування методів багатовимірною аналізу, загальні принципи та можливості яких стисло описані нижче.

**Покрокова багатовимірна регресія** дозволяє встановити рівняння лінійної залежності однієї з ознак від декількох інших. Багатофакторні залежності та методика їх побудови за методом найменших квадратів описані в підрозділ 4.3. Як і в описаному випадку, параметри лінійної апроксимуючої функції визначаються методом найменших квадратів [2] з умови мінімізації суми квадратів відхилень дослідних значень залежної ознаки від відповідних значень апроксимуючої функції. Суть покрокової регресії полягає у застосуванні спеціального алгоритму, який на кожному кроці його виконання додає в рівняння регресії одну незалежну змінну, тобто один стовпець матриці (5.10). Для цього обирається саме та змінна, яка на даному кроці розрахунку забезпечить найістотніше зменшення суми квадратів відхилень. Включення кожної наступної змінної підвищує точність апроксимації, але ускладнює отриману формулу залежності. Завдяки покроковому оцінюванню точності можна обрати компромісне рішення між точністю та складністю апроксимуючої функції. При необхідності отримання нелінійної залежності від певної ознаки слід ввести додаткову ознаку, обчислену шляхом необхідного нелінійного перетворення обраної ознаки (піднесення до ступеню, логарифмування чи інших дій) з наступним використанням алгоритму покрокової регресії.

**Метод головних компонент** дозволяє згрупувати ознаки  $x_{ij}$  у декілька незалежних один від одного узагальнених факторів  $y_i$ , кожен з яких узагальнює декілька ознак і визначається за формулою

$$y_i = \sum_j a_j x_{i,j} . \quad (5.11)$$

Параметри  $a_j$  функцій (5.11) визначаються таким чином, щоб узагальнені фактори  $y_i$  були незалежними один від одного. Фактори обирають таким чином, щоб вони групували початкові ознаки за якимось принципом: фактор розмірів, фактор фізичних властивостей, фактор хімічного складу тощо. Метод головних компонент дозволяє зменшити кількість незалежних змінних, якими характеризуються об'єкти дослідження, за рахунок заміни початкових ознак  $x_{ij}$  з матриці даних (5.10) набагато меншою кількістю узагальнених факторів  $y_i$ .

**Метод канонічних кореляцій** полягає в тому, що з наявних ознак дослідник виділяє дві групи, які відображають два комплекси властивостей об'єктів (наприклад, фізичні властивості та хімічний склад). Виділені групи не обов'язково повинні охоплювати всі наявні ознаки. Наприклад, при наявності 10 ознак, тобто 10 стовпців матриці (5.10), можна включити до виділених груп лише 7 ознак:

група 1 –  $x_{i,1}, x_{i,3}, x_{i,4}, x_{i,8}$ ; група 2 –  $x_{i,2}, x_{i,7}, x_{i,9}$ .

За даними кожної з груп програма будує лінійні функції виду

$$\begin{aligned} y_i &= a_1 x_{i,1} + a_3 x_{i,3} + a_4 x_{i,4} + a_8 x_{i,8}; \\ z_i &= a_2 x_{i,2} + a_7 x_{i,7} + a_9 x_{i,9}. \end{aligned} \quad (5.12)$$

Коефіцієнти  $a_j$  підбираються таким чином, щоб корельованість отриманих нових змінних  $y$  і  $z$  була якомога вищою. Нові змінні  $y$  і  $z$  узагальнено відображають обрані комплекси ознак об'єктів аналізу, а рівняння лінійної регресії та коефіцієнт кореляції між  $y$  і  $z$  відображають залежність між обраними комплексами ознак (наприклад, між фізичними властивостями та хімічним складом об'єктів дослідження). Результатом такого аналізу є узагальнений висновок щодо наявності та тісноти залежності між двома комплексами ознак об'єктів дослідження. Наприклад, можна зробити висновок, що фізичні властивості та хімічний склад об'єктів дослідження є практично незалежними, або навпаки – тісно пов'язаними між собою.

**Метод дискримінантних функцій** дозволяє з урахуванням наявного комплексу ознак, представлених для кожного об'єкта рядком матриці (5.10), здійснити класифікацію цих об'єктів на дві групи. Наприклад, за результатами вимірювань, спостережень чи аналізів можна класифікувати: вироби – на якісні та неякісні, природні явища – на безпечні й небезпечні, кліматичні умови – на комфортні та некомфортні, людей – на здорових і хворих. Спочатку формується навчаюча вибірка об'єктів, у якій для кожного об'єкта відомі значення ознак  $x_{i,j}$  та його належність до одного з двох класів. За навчаючою вибіркою будується лінійна дискримінантна функція виду

$$D = \sum_j a_j x_{i,j}, \quad (5.13)$$

коефіцієнти якої  $a_j$  визначаються таким чином, щоб значення (5.13) для обох класів відрізнялися якомога сильніше. Встановлюється граничне значення дискримінантної функції  $D_0$ , яке з достатньо високою впевненістю розділяє об'єкти з навчаючої вибірки на два класи.

Для класифікації (діагностики) нового  $i$ -того об'єкта за комплексом його ознак  $x_{i,j}$  обчислюється дискримінантна функція (5.13). Залежно від отриманого значення дискримінантної функції  $D > D_0$  чи  $D < D_0$  об'єкт відноситься до одного з класів. Достовірність класифікації (діагностики) нових об'єктів залежить від повноти наявного набору ознак (5.10) та обсягу навчаючої вибірки. Недостатній перелік ознак (5.10) не дозволить побудувати дискримінантну функцію (5.13), яка досить чітко розділяє об'єкти аналізу на дві групи, а це збільшить імовірність неправильної класифікації (діагностики) нових об'єктів.

### 5.3. Функції випадкових величин

При розв'язанні науково-технічних задач доводиться оперувати з функціями, аргументами яких є випадкові величини. Результати перетворення також є випадковими величинами з власними законами розподілу та числовими характеристиками. Так наприклад, перехід від вимірної на метеостанції швидкості вітру  $v$  до вітрового тиску на будівлю  $w$  здійснюється за формулою

$$w = 0,61 v^2, \quad (5.14)$$

а зусилля  $Y$  в елементі несучої будівельної конструкції, металевій конструкції вантажопідйомного крана чи дорожньої машини є лінійною комбінацією навантажень  $X_i$ :

$$Y = \sum_{i=1}^N \alpha_i X_i, \quad (5.15)$$

де  $\alpha_i$  – коефіцієнт впливу  $i$ -го навантаження з  $N$  діючих на конструкцію.

Навіть найпростіші нелінійні перетворення подібні до (5.14) є досить складними задачами, бо вони можуть принципово змінювати не лише числові характеристики, але й форми законів розподілу випадкових величин. Загальна методика визначення законів розподілу та числових характеристик нелінійних функцій випадкових величин з довільними законами розподілу викладена в [2].

Найпростішими є операції з константами, описані в підрозділі 2.5. Наведені там формули дозволяють визначити математичне сподівання та стандарт випадкової величини, отриманої у результаті додавання чи множення вихідної величини на константу. Форма закону розподілу у результаті виконання таких операцій не змінюється.

Досить детально розроблений математичний апарат *лінійних перетворень* випадкових величин, відомий в теорії імовірностей [2] як *теорема про числові характеристики*. Математичне сподівання та стандарт лінійної комбінації (5.15)  $N$  випадкових величин  $X_i$  визначаються за формулами

$$M_y = \sum_{i=1}^N \alpha_i M_i, \quad S_y = \sqrt{\sum_{i=1}^N (\alpha_i S_i)^2}, \quad (5.16)$$

де  $M_i$  та  $S_i$  – математичне сподівання та стандарт  $i$ -тої складової.

Згідно з формулами (5.16), математичне сподівання функції (5.15) дорівнює алгебраїчній сумі математичних сподівань складових. Дисперсії, рівні квадратам стандартів  $S_i^2$ , завжди додаються, незалежно від знаків множників  $\alpha_i$ , тому дисперсія різниці випадкових величин також дорівнює сумі дисперсій складових.

Додавання чи віднімання декількох випадкових величин згідно з формулами (5.15) та (5.16) змінює не тільки числові характеристики, але й вид закону розподілу результуючої випадкової величини. Стійким до операції додавання є нормальний розподіл, тому лінійна комбінація (5.15) нормально розподілених випадкових величин також має нормальний розподіл. Ця властивість нормального розподілу часто використовується при розв'язанні різних науково-технічних задач.

При інших законах розподілу, нестійких до лінійних перетворень, а також при нелінійних перетвореннях випадкових величин закон розподілу результату може бути визначений загальними методами, викладеними в [2]. У більшості випадків неможливо отримати замкнуте аналітичне рішення за цією методикою. Це обумовлює необхідність чисельного інтегрування вихідних розподілів та замість загального рішення у вигляді формули дає лише конкретний числовий результат для конкретних значень параметрів вихідних розподілів.

Одним з методів наближеного визначення числових характеристик довільних функцій однієї чи декількох випадкових величин є *метод лінеаризації*, детально описаний в підручнику [2]. Суть цього методу полягає в тому, що в околі математичних сподівань аргументів реальна нелінійна функція замінюється лінійною, згідно з якою й визначаються числові характеристики результуючої випадкової величини на основі формул (5.16). В якості лінійних наближень по кожному з аргументів використовуються дотичні до реально заданої нелінійної функції.

Сформулюємо загальну постановку задачі та принципи визначення числових характеристик нелінійної функції декількох випадкових величин методом лінеаризації згідно з вказівками [2]. Необхідно визначити числові характеристики функції взаємно незалежних випадкових аргументів  $X_i$ ,  $i = 1 \dots N$ , яка має загальний вигляд:

$$Y = \varphi(X_1, X_2, X_i \dots X_N). \quad (5.17)$$

Згідно з вказівками [2], математичне сподівання випадкової величини  $Y$  дорівнює

$$M_Y = \varphi(M_1, M_2, M_i \dots M_N), \quad (5.18)$$

де  $M_i$  – математичні сподівання аргументів  $X_i$ .

Дисперсія випадкової величини  $Y$  визначається за формулою

$$D_Y = \sum_{i=1}^N \left( \frac{\partial \varphi}{\partial X_i} \right)^2 D_i, \quad (5.19)$$

де  $D_i$  – дисперсії аргументів  $X_i$ ;

$\frac{\partial \varphi}{\partial X_i}$  – частинні похідні функції (5.17) по усіх аргументах  $X_i$ .

Для розв'язання конкретної задачі з нелінійного перетворення випадкових величин необхідно:

- визначити математичні сподівання  $M_i$  та дисперсії  $D_i = S_i^2$  усіх аргументів  $X_i$ ;
- записати конкретний вираз функції (5.17), взяти її похідні по всіх аргументах та записати ці похідні для значень аргументів, рівних їх математичним сподіванням, тобто при  $X_i = M_i$ ;
- підставити отримані вирази похідних до (5.19) та виконати необхідні алгебраїчні перетворення;
- в отримані робочі формули підставити відомі значення  $M_i$  та  $D_i$  для усіх  $N$  аргументів та обчислити шукані значення математичного сподівання  $M_Y$  та дисперсії  $D_Y$  результуючої випадкової величини  $Y$ .

Якщо потрібно знайти числові характеристики результату подібного до (5.14) чи іншого нелінійного перетворення однієї випадкової величини, у формулах (5.17), (5.18) і (5.19) замість декількох аргументів  $X_i$  залишається лише один аргумент. При цьому загальна процедура розв'язку не змінюється, але робочі формули спрощуються за рахунок відкидання усіх аргументів окрім  $X_j$ .

Додатково відмітимо, що метод лінеаризації дозволяє визначити математичне сподівання, дисперсію та стандарт випадкової величини, отриманої у результаті нелінійного перетворення одного чи декількох випадкових аргументів. Закон розподілу результуючої випадкової величини залишається невизначеним.

Складні нелінійні перетворення випадкових величин можна також виконувати методом статистичного моделювання, який також називають *методом Монте-Карло*. Суть цього універсального методу полягає в тому, що за допомогою спеціальних комп'ютерних програм моделюються вибірки вхідних величин достатньо великого обсягу (тисячі чи навіть сотні тисяч значень). Для кожного елемента змодельованої вибірки (або сукупності елементів декількох вибірок) виконується необхідне нелінійне перетворення типу (5.17) шляхом простого обчислення за відомою формулою чи алгоритмом перетворення. Таким чином у результаті тисяч моделювань формується вибірка результатів нелінійного перетворення однієї чи декількох випадкових величин з довільними законами розподілу. Статистична обробка результуючої вибірки дозволяє отримати числові характеристики та закон розподілу результуючої випадкової величини.

До переваг методу Монте-Карло слід віднести його універсальність, простоту реалізації, а також можливість отримання не лише числових характеристик, але й закону розподілу результату. Основним недоліком цього методу є неможливість отримання розв'язку в загальному вигляді. Отримані в результаті моделювання значення числових характеристик та закон розподілу результуючої випадкової величини відповідають заданим числовим значенням параметрів законів розподілу аргументів.

#### **5.4. Поняття про випадкові процеси**

Імовірнісна модель випадкової величини використовується для опису величин, незмінних у часі. У практиці наукових досліджень часто доводиться оперувати з величинами, які випадковим чином змінюються в часі. Такими є переважна більшість змінних навантажень і впливів на будівлі, будівельні конструкції, машини і механізми, процеси накопичення експлуатаційних пошкоджень, зміни технічних характеристик матеріалів унаслідок старіння та деградації, а також інші подібні явища.

**Випадковими процесами** називають функції часу, які змінюються випадковим чином і значення яких неможливо передбачити до проведення досліду чи спостереження. Значення випадкового процесу в фіксований момент часу називають його ординатою, а послідовність цих значень, отриману в результаті спостереження протягом певного відрізка часу – реалізацією випадкового процесу.

Випадкові процеси характеризуються законом розподілу ординати та його числовими характеристиками, а також характеристиками частотної структури, які відображають швидкість випадкових змін ординати. **Стаціонарними** називають випадкові процеси, характеристики яких є незмінними в часі. Процеси, числові характеристики яких істотно змінюються в часі, називають нестаціонарними. В якості прикладу на рисунку 5.1 наведені експериментально отримані реалізації випадкових процесів змін температури атмосферного повітря та повітря в житловому приміщенні на протязі опалювального періоду.

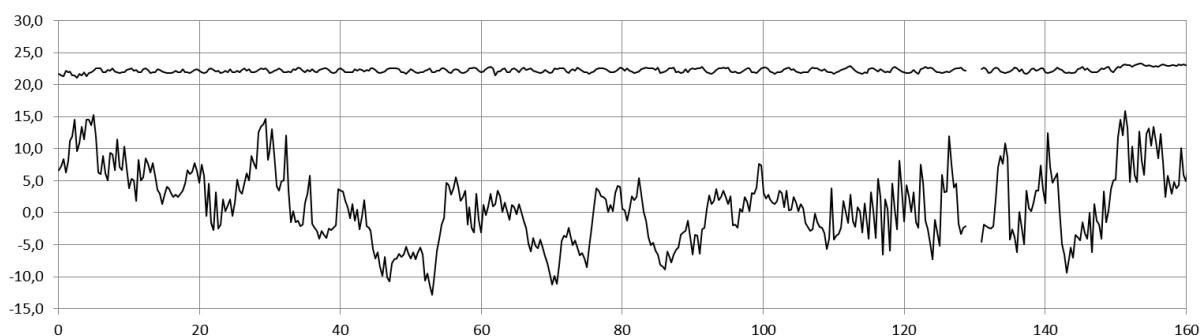


Рис. 5.1. Реалізації випадкових процесів змін температури повітря

З рисунка видно, що зміни температури повітря в приміщенні утворюють стаціонарний випадковий процес із середнім значенням близько  $22^{\circ}\text{C}$  та розмахом в межах  $2^{\circ}\text{C}$ . Процес змін температури атмосферного повітря (нижня реалізація на рисунку 5.1) є нестаціонарним унаслідок виражених сезонних змін погоди. Однак на невеликому відрізку часу порядку тижня чи декади зміни середнього значення й стандарту є незначними, що дозволяє приблизно вважати цей процес стаціонарним на таких відрізках часу. Такого виду випадкові процеси, числові характеристики яких змінюються набагато повільніше порівняно з випадковими змінами ординати, називають **квазістаціонарними**. Відомо [3], що імовірнісною моделлю квазістаціонарного випадкового процесу добре описуються більшість кліматичних навантажень і впливів на будівельні конструкції, будівлі та споруди.

Не дивлячись на непередбачуваність змін, близькі за часом ординати випадкових процесів пов'язані між собою, а збільшення різниці в часі призводить до послаблення такого зв'язку. Наприклад, температура повітря чи швидкість вітру протягом години змінюються мало, протягом доби можуть відбутися значні зміни, а ординати, віддалені на декілька діб, є практично повністю незалежними. Міра зв'язності випадкового процесу характеризується його *нормованою кореляційною функцією*, яка близька до одиниці для сусідніх ординат і зменшується по мірі їх віддалення в часі. Більш простим параметром частотної структури випадкового процесу є *ефективна частота*, рівна середньозваженій частоті періодичних змін ординати. Ефективна частота відображає швидкість випадкових коливань дослідженого процесу і використовується при розв'язанні більшості задач дослідження й нормування навантажень на будівельні конструкції.

Більш детальну інформацію про випадкові процеси, зокрема їх класифікацію, можливості використання для імовірнісного опису роботи будівельних конструкцій та машин, способи визначення статистичних характеристик, методи оперування при розв'язанні науково-інженерних задач можна знайти в літературі [2, 3]. Розроблення імовірнісних моделей випадкових процесів, включаючи визначення необхідних статистичних характеристик, вимагає виконання досить складного аналізу статистичних даних, обсяг яких багатократно перевищує звичайні обсяги вибірок випадкових величин. Набагато складнішим є також оперування з імовірнісними моделями випадкових процесів при розв'язанні наукових та інженерних задач, що істотно обмежує галузь використання цих моделей.

З прикладами практичного застосування моделей випадкових процесів, зокрема квазістаціонарного диференційованого процесу, для розв'язання задач дослідження й нормування кліматичних навантажень на будівельні конструкції можна ознайомитися в монографії [3].

### **5.5. Математичне забезпечення багатовимірної статистичного аналізу в середовищі Microsoft Excel**

В наведеній нижче таблиці надано перелік, призначення та короткий опис параметрів функцій Microsoft Excel, які використовуються для порівняння випадкових величин та їх статистичних характеристик. Як і раніше, аргумент "масив" означає сукупність комірок у вигляді стовпчика чи рядка, який містить необхідні аргументи чи функції.

Запис функції	Призначення функції та опис параметрів
<i>КОРРЕЛ(масив_X; масив_Y)</i>	Обчислює коефіцієнт кореляції двох вибірок випадкових величин, заданих полями комірок "масив_X" і "масив_Y"
<i>СУММКВ(масив_X)</i>	Обчислює суму квадратів чисел, заданих у полі комірок "масив_X"
<i>СЧЕТ(список)</i>	Підраховує обсяг вибірки, заданої в полі комірок "список"
<i>СРЗНАЧ(список)</i>	Обчислює середнє значення вибірки, заданої в полі комірок "список"
<i>СТАНДОТКЛОН(список), СТАНДОТКЛОН.В(список)</i>	Обчислює стандарт вибірки, заданої в полі комірок "список"
<i>НОРМ.СТ.ОБР(1 – <math>\alpha</math> / 2)</i>	Обчислює двостороннє критичне значення для порівняння математичних сподівань двох випадкових величин
<i>F.ОБР.ПХ(<math>\alpha; k_1; k_2</math>)</i>	Обчислює правостороннє критичне значення критерія Фішера для порівняння дисперсій двох випадкових величин
<i>Розділ меню Excel «Дані» → «Аналіз даних» → «Однофакторний дисперсійний аналіз»</i>	Виконує дисперсійний аналіз для декількох вибірок на заданому рівні значимості

## 6. ПРИКЛАДИ ЗАСТОСУВАННЯ СТАТИСТИЧНИХ МЕТОДІВ ПРИ ВИКОНАННІ ІНЖЕНЕРНИХ ДОСЛІДЖЕНЬ

В даному розділі наведені приклади розв'язання задач статистичного аналізу, які часто зустрічаються в практиці інженерних досліджень та методики розв'язання яких описані в попередніх розділах. По кожному з прикладів надана постановка задачі, посилання на викладену в попередніх розділах методику її розв'язання, а також копії (скрін-шоти) робочих аркушів Microsoft Excel та пояснення змісту окремих комірок. Ці матеріали разом з відповідними робочими аркушами файлу "Аналіз\_даних.xlsx", який додається до даного посібника, дозволяють детально вивчити можливі схеми організації обчислень та розробити аналогічні розрахункові бланки для розв'язання власних задач. В усіх прикладах імена робочих аркушів вказаного файлу Excel відповідають номерам наведених нижче підрозділів, у яких викладені розв'язки відповідних задач.

### 6.1. Аналіз середньої густини пінополістиролу

Для дослідження використані результатами випробувань 32 зразків плит пінополістиролу товщиною 40 мм із заявленою маркою за середньою густиною D 25. Середня густина усіх зразків визначена за стандартною методикою ДСТУ Б В.2.7-38-95 "Матеріали і вироби теплоізоляційні. Методи випробувань", виражена в  $\text{кг}/\text{м}^3$  і наведена в комірках В 5...В 36 робочого аркуша "6.1" файлу "Аналіз\_даних.xlsx", копія якого зображена на рисунку 6.1. Статистична обробка наявної вибірки значень середньої густини виконана за методикою підрозділів 3.1, 3.2, 3.3. Імовірнісне подання випадкової величини середньої густини пінополістиролу нормальним законом розподілу та перевірка узгодженості цього закону з дослідними даними здійснені за вказівками підрозділів 2.4.2 і 3.5.

З рисунка 6.1 видно, що густина випробуваних зразків змінюється від  $15,48 \text{ кг}/\text{м}^3$  до  $17,75 \text{ кг}/\text{м}^3$  при середньому значенні  $M = 16,58 \text{ кг}/\text{м}^3$ . Під стовпчиком даних з густини у комірках В 38...В 42 записані функції для визначення числових характеристик вибірки: середнього значення  $M$ , стандарту  $S$ , коефіцієнта варіації  $V$  та коефіцієнта асиметрії  $A$ . Коефіцієнт варіації  $V = 0,032$  вказує на високу однорідність випробуваного матеріалу, а невеликий коефіцієнт асиметрії  $A = 0,332$  на близькість розподілу середньої густини до симетричного.

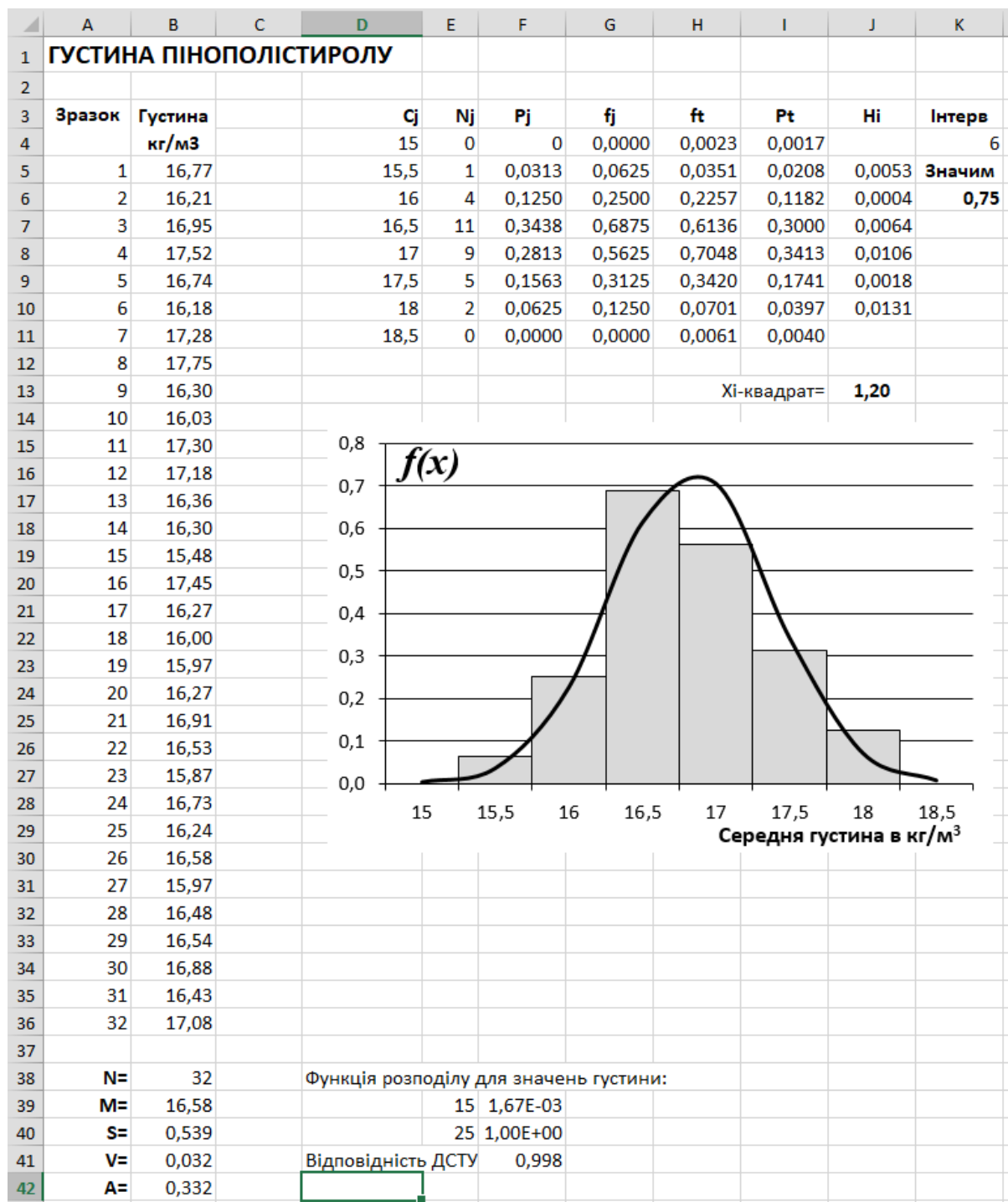


Рис. 6.1. Статистична обробки даних по густині пінополістиролу

Побудова гістограми розподілу реалізована в комірках D 3...G 11. Столпчики цього діапазону містять задані кінці інтервалів гістограми  $C_j$ , підраховані функцією  $ЧАСТОТА(...)$  кількості даних в інтервалах  $N_j$ , імовірності попадання до інтервалів  $P_j$  за формулою (3.2) та дослідні значення густини розподілу  $f_j$  за формулою (3.3). За рекомендацією таблиці 3.1 для обсягу вибірки  $N = 32$  зразки встановлено 6 інтервалів гістограми з кроком  $d = 0,5 \text{ кг/м}^3$  в межах від  $15 \text{ кг/м}^3$  до  $18 \text{ кг/м}^3$ . Графічне

зображення гістограми розподілу, наведене на рисунку 6.1, вказує на можливість використання для імовірнісного опису середньої густини пінополістиролу нормального закону розподілу.

Комірки діапазону Н 3...J 13 містять результати побудови кривої густини нормального розподілу  $f_t$  за функцією *НОРМРАСП(...)*, обчислення теоретичних імовірностей попадання до інтервалів  $P_t$  як різниць функцій *НОРМРАСП(...)* для кінця й початку кожного інтервалу за (3.20), а також значення статистики критерію Пірсона  $H_i$  в кожному інтервалі. Сумарне значення статистики критерію Пірсона (3.19) визначене в комірці J 13.

З урахуванням отриманого значення статистики  $\chi^2 = 1,20$  та кількості ступенів свободи  $k = 6 - 3 = 3$  (кількість інтервалів гістограми, зменшена на три) за таблицею А.2 отримуємо рівень значимості  $\alpha > 0,20$ . В комірці К 6 за функцією *ХИ2.РАСП(...)* встановлена точна величина рівня значимості  $\alpha = 0,75$ . Оскільки  $\alpha > 0,05$ , нормальний закон розподілу не суперечить дослідній гістограмі і може використовуватися для імовірнісного опису середньої густини пінополістиролу.

Згідно з вимогами ДСТУ Б В.2.7-8:94 "Плити пінополістирольні. Технічні умови", середня густина пінополістиролу марки 25 повинна знаходитися в межах від  $15 \text{ кг/м}^3$  до  $25 \text{ кг/м}^3$ . З використанням функції нормального розподілу *НОРМРАСП(...)* в комірках F 39 та F 40 визначені імовірності того, що середня густина випробуваного пінополістиролу буде меншою за  $15 \text{ кг/м}^3$  та  $25 \text{ кг/м}^3$ . Наведена в комірці F 40 різниця цих значень є імовірністю того, що випробуваний пінополістирол належить до заявленої марки 25 за середньою густиною. Імовірність 0,998 вказує, що лише два зразки з тисячі можуть не відповідати марці 25 за середньою густиною. Така висока забезпеченість дозволяє стверджувати, що середня густина випробуваного пінополістиролу повністю відповідає вимогам стандарту до марки 25.

## 6.2. Дослідження та нормування снігового навантаження

Для дослідження використані результати снігомірних зйомок, які проводилися на метеостанції м. Кропивницький у період з 1950 року по 2017 рік. З наявних даних сформована вибірка із  $N = 67$  значень річних максимумів ваги снігового покриву, які занесені в комірки В 4...К 10 робочого аркуша "6.2" з файлу Microsoft Excel "Аналіз\_даних.xlsx", копія якого зображена на рисунку 6.2.

	A	B	C	D	E	F	G	H	I	J	K
1	<b>ДОСЛІДЖЕННЯ ВАГИ СНІГОВОГО ПОКРИВУ в Кропивницькому</b>										
2											
3	<b>Роки</b>	<b>Річні максимуми ваги снігового покриву в Па</b>									
4	1950-1960	307	448	710	290	270	490	170	220	24	420
5	1960-1970	354	240	670	750	480	990	960	760	590	360
6	1970-1980	550	150	490	270	142	220	360	210	448	189
7	1980-1990	400	330	310	200	600	470	730	280	140	170
8	1990-2000	100	47	700	94	80	1380	230	378	400	380
9	2000-2010	340	120	800	840	650	410	80	180	260	990
10	2010-2017	360	620	200	190	140	810	520			
11											
12	<b>Характеристики</b>		<b>Гістограма розподілу</b>				<b>Розподіл Гумбеля</b>				
13	<b>N=</b>	<b>67</b>	<b>C<sub>j</sub></b>	<b>N<sub>j</sub></b>	<b>P<sub>j</sub></b>	<b>f<sub>j</sub></b>	<b>f</b>	<b>F<sub>t</sub></b>	<b>P<sub>t</sub></b>	<b>Хі-квадр</b>	
14	Min	23,6	0				0,00028	0,0343			
15	Max	1380	100	6	0,0896	0,00090	0,00078	0,1120	0,0777	0,0018	
16	M=	410	200	12	0,1791	0,00179	0,00130	0,2413	0,1293	0,0192	
17	S=	274	300	10	0,1493	0,00149	0,00157	0,3972	0,1559	0,0003	
18	V=	0,668	400	12	0,1791	0,00179	0,00153	0,5491	0,1518	0,0049	
19	A=	1,089	500	8	0,1194	0,00119	0,00129	0,6775	0,1284	0,0006	
20	Alfa	282	600	4	0,0597	0,00060	0,00099	0,7766	0,0991	0,0157	
21	Beta	232	700	4	0,0597	0,00060	0,00072	0,8486	0,0720	0,0021	
22			800	5	0,0746	0,00075	0,00050	0,8989	0,0503	0,0118	
23	<b>T</b>	<b>Q(T)</b>	900	2	0,0299	0,00030	0,00034	0,9331	0,0342	0,0006	
24	10	803	1000	3	0,0448	0,00045	0,00023	0,9560	0,0229	0,0208	
25	20	969	1100	0	0,0000	0,00000	0,00015	0,9712	0,0152	0,0152	
26	30	1065	1200	0	0,0000	0,00000	0,00010	0,9812	0,0100	0,0100	
27	50	1185	1300	0	0,0000	0,00000	0,00006	0,9878	0,0065	0,0065	
28	70	1264	1400	1	0,0149	0,00015	0,00004	0,9920	0,0043	0,0265	
29	100	1347	1500	0	0,0000	0,00000	0,00003	0,9948	0,0028	0,0028	
30	150	1441									
31	200	1508							<b>Хі-квадрат=</b>	<b>9,30</b>	
32											

Рис. 6.2. Аналіз ваги снігового покриву в м. Кропивницький

Числові характеристики вибірки визначені в комірках В 13...В 19 з використанням функцій, описаних у підрозділі 3.6. Комірки В 20 та В 21 містять значення параметрів розподілу Гумбеля (2.26).

Розрахунки для побудови гістограми розподілу виконані в комірках D 13...G 29 таким же чином, як і в попередньому прикладі. За заданими межами інтервалів гістограми  $C_j$ , та вибіркою значень ваги снігового покриву з комірок В 4...К 10 обчислені: кількості даних в інтервалах  $N_j$  за функцією *ЧАСТОТА(...)*, імовірності попадання до інтервалів  $P_j$  за формулою (3.2) та дослідні значення густини розподілу  $f_j$  за (3.3).

Хоча таблиця 3.1 для обсягу вибірки  $N = 40...100$  рекомендує 5...8 інтервалів, при округленій довжині інтервалу  $d = 100$  Па отримана гістограма розподілу з 14 інтервалів. Це обумовлено асиметричністю гістограми, права частина якої містить три "пустих" інтервали та одне значення в інтервалі 1300...1400 Па, яке відображає екстремальне

значення ваги снігового покриву 1380 Па зимою 1995-1996 року. Окрім того, збільшення довжини інтервалу не дасть змоги виявити форму розподілу при малих значеннях ваги снігового покриву.

Графічне зображення гістограми розподілу, наведене на рисунку 6.3. Вона має явно асиметричний характер, подібний до кривої розподілу Гумбеля, який зазвичай використовується для імовірнісного опису річних максимумів ваги снігового покриву [3]. На близькість до розподілу Гумбеля вказує також обчислений в комірці В 19 коефіцієнт асиметрії  $A = 1,089$  (у підрозділі 2.4.4 вказано, що для розподілу Гумбеля  $A = 1,14$ ).

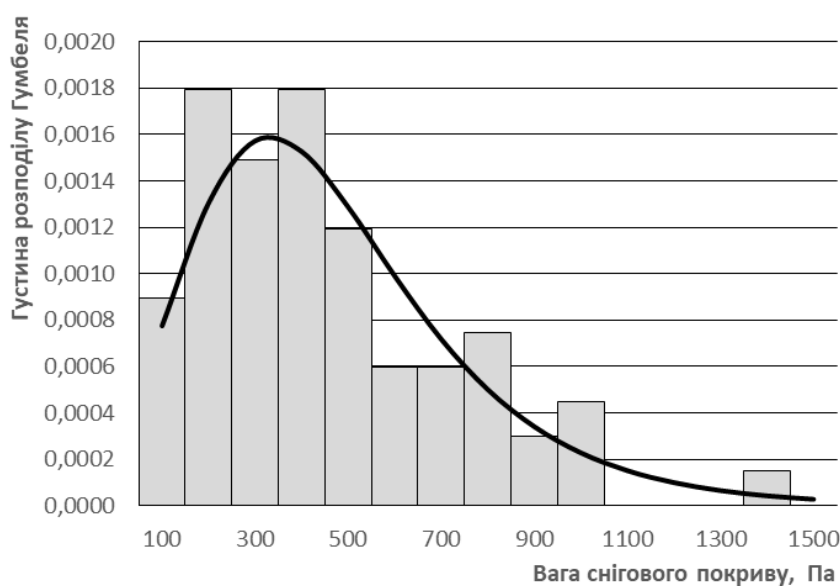


Рис. 6.3. Гістограма розподілу річних максимумів ваги снігового покриву

Крива густини розподілу Гумбеля (2.23), побудована на рисунку 6.3 за даними комірок Н 14...Н 29, візуально відповідає гістограмі розподілу. Перевірка відповідності закону розподілу Гумбеля до дослідних даних здійснена в комірках І 13...К 31 робочого аркуша Excel, зображеного на рисунку 6.2. Там обчислені значення інтегральної функції розподілу Гумбеля (2.24), теоретичні імовірностей попадання ваги снігового покриву до усіх інтервалів  $P_i$ , а також значення статистики критерію Пірсона  $H_i$  в кожному інтервалі. Сумарне значення статистики критерію Пірсона (3.19)  $\chi^2 = 9,30$  наведене в комірці К 31.

При кількості ступенів свободи  $k = 14 - 3 = 11$  (кількість інтервалів гістограми, зменшена на три) та рівні значимості  $\alpha = 0,1$  за таблицею А.2 отримуємо критичне значення статистики критерію Пірсона  $\chi^2_{cr} = 17,3$ . Отже закон розподілу Гумбеля не суперечить дослідним даним і може бути використаний для імовірнісного опису річних максимумів ваги снігового покриву в м. Кропивницький.

Отриманий закон розподілу дозволяє встановити розрахункові значення снігового навантаження на поверхню ґрунту для різних періодів повторюваності та характеристичне значення снігового навантаження. Згідно з вимогами ДБН В.1.2-2:2006 "Навантаження і впливи" та відомою методикою нормування снігового навантаження [3], граничні розрахункові значення встановлюють залежно від середнього періоду повторюваності, а характеристичне значення може перевищуватися в середньому один раз на 50 років, тобто має період повторюваності  $T = 50$  років.

Виразивши у формулі (2.27) для визначення квантилів розподілу Гумбеля імовірність  $P$  через період повторюваності  $T$ , отримуємо робочу формулу для обчислення граничних розрахункових значень навантаження

$$Q(T) = \alpha - \beta \times \ln \left[ -\ln \left( 1 - \frac{1}{T} \right) \right], \quad (6.1)$$

за якою в комірках DB 24...B 31 робочого аркуша з рисунка 6.2 обчислені граничні розрахункові значення для декількох періодів повторюваності в межах від 10 до 200 років.

Характеристичне значення ваги снігового покриву  $Q(50) = 1185$  Па отримане на 4% меншим, ніж наведене в ДБН В.1.2-2:2006 значення для м. Кропивницький 1230 Па.

### 6.3. Аналіз даних по швидкості вітру

Дослідження виконане з метою прогнозування імовірного часу простоїв монтажних кранів у результаті дії сильних вітрів в районі м. Кропивницький. Використані для аналізу дані метеорологічних щомісячників містять таблиці з частотами реалізації різних швидкостей вітру, тобто табличне подання гістограм розподілу. Узагальнивши дані за 20 років спостереження, отримуємо гістограму, представлену таблицею в комірках A 3...F 17 робочого аркуша робочого аркуша "6.3" з файлу Microsoft Excel "Аналіз\_даних.xlsx", зображення якого наведене на рисунку 6.4. Значення  $C_j$  за (3.8) дорівнюють серединам інтервалів, межі яких вказані в стовпцях A і B. Кількості даних в кожному інтервалі  $N_j$  задані в метеорологічних щомісячниках, а імовірності попадання в інтервали  $P_j$  та значення густини розподілу  $f_j$  обчислені за формулами (3.2) і (3.3). Графічне зображення гістограми розподілу швидкості вітру наведене на рисунку 6.5.

	A	B	C	D	E	F	G	H	I
1	<b>АНАЛІЗ ШВИДКОСТІ ВІТРУ в Кропивницькому</b>								
2	Інтервал швидкості вітру			Гістограма розподілу					Теорет.
3	початок	кінць	Cj	Nj	Pj	fj	(Cj-M)^2	(Cj-M)^3	f(Cj)
4	0	1	0,5	7830	0,13095	0,08730	12,07	-41,908	0,060971
5	2	3	2,5	16666	0,27871	0,13936	2,17	-3,199	0,179828
6	4	5	4,5	23491	0,39285	0,19643	0,28	0,146	0,152929
7	6	7	6,5	8107	0,13558	0,06779	6,38	16,127	0,077511
8	8	9	8,5	2879	0,04815	0,02407	20,49	92,745	0,026414
9	10	11	10,5	622	0,01040	0,00520	42,60	278,000	0,006348
10	12	13	12,5	162	0,00271	0,00135	72,70	619,890	0,001104
11	14	15	14,5	30	0,00050	0,00025	110,81	1166,417	0,000141
12	16	17	16,5	8	0,00013	0,00007	156,91	1965,580	0,000013
13	18	20	19	1	0,00002	0,00001	225,80	3392,929	0,000000
14	21	24	22,5	0	0,00000	0,00000	343,23	6358,888	0,000000
15	25	28	26,5	0	0,00000	0,00000	507,44	11430,942	0,000000
16	29	34	31,5	0	0,00000	0,00000	757,71	20857,089	0,000000
17	35	40	37,5	0	0,00000	0,00000	1124,03	37684,716	0,000000
18									
19			N=	59796			Alfa	1,8708	
20			M=	3,97			Beta	0,0606	
21			S=	2,21			Цільова	9,036E-09	
22			V=	0,555		Тривалість v>10 м/с		97,3	
23			A=	0,541		Тривалість v>15 м/с		0,59	

Рис. 5.4. Статистичний аналіз швидкості вітру в м. Кропивницький

Числові характеристики розподілу швидкості вітру визначені в комірках D 20...D 23 за формулами (3.9), для чого у стовпцях G, H обчислені квадрати та куби відхилень середин інтервалів від середнього значення  $M$  з комірки D 20.

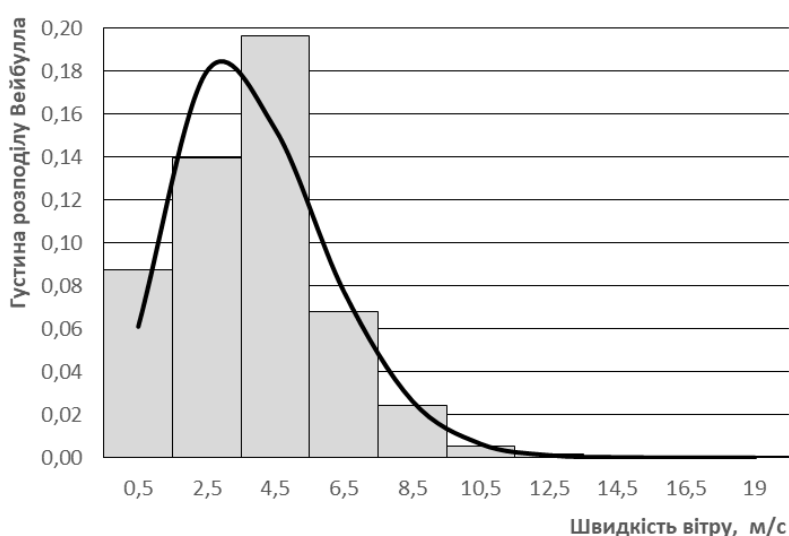


Рис. 6.5. Гістограма розподілу швидкості вітру в м. Кропивницький

Відомо [3], що поточні значення швидкості вітру, добре описуються законом розподілу Вейбулла (2.28), (2.29). Параметр  $\alpha$  визначено шляхом розв'язання нелінійного рівняння (2.30) з використанням інструменту Microsoft Excel "поиск решения". Для цього в комірці Н 21 записано вираз для цільової функції пошуку (2.30), після чого розв'язана задача: встановити значення комірки Н 21 рівним нулю, змінюючи комірку Н 19. Параметр  $\beta$  обчислений в комірці Н 20 за формулою (2.31).

За отриманими значеннями параметрів  $\alpha$  і  $\beta$  в комірках І 4...І 17 обчислена густина розподілу Вейбулла (2.28), яка зображена на гістограмі з рисунка 6.5. Візуально ця крива добре відповідає дослідній гістограмі. Перевірка узгодженості не виконувалася, оскільки критерій Пірсона має на увазі незалежність значень у вибірці, а наявні результати строкових вимірювань (кожні 6 годин) є взаємно залежними.

Отриманий у результаті виконаної статистичної обробки розподіл Вейбулла дозволяє визначити імовірність перевищення заданих значень швидкості вітру за формулою інтегральної функції розподілу (2.29). Перемноживши значення  $1 - F(x)$  на тривалість року  $365 \times 24 = 8760$  годин, отримуємо робочу формулу для визначення часу, протягом якого в середньому може перевищуватися задана швидкість вітру  $v_m$ :

$$T(v_m) = 8760 \times \exp(-\beta x^\alpha). \quad (6.2)$$

Правила технічної експлуатації вантажопідйомних кранів вимагають припинення робіт при швидкості вітру  $v_m = 15$  м/с (при таких швидкостях вітру оголошується штормове попередження), а правилами техніки безпеки при виконанні будівельно-монтажних робіт заборонено вести монтаж конструкцій з великою парусністю при швидкостях вітру понад 10 м/с. В комірках Н 22 і Н 23 робочого аркуша з рисунка 6.4 за формулою (6.2) обчислені імовірні тривалості перевищень вказаних гранично допустимих швидкостей вітру.

Результати розрахунків показують, що в умовах м. Кропивницький швидкість вітру  $v_m = 15$  м/сек перевищується в середньому протягом 0,6 години протягом року. Отже випадки, коли кран необхідно повністю зупиняти і ставити на гальма, є мало імовірними й загалом не тривалими. Тривалість перевищення швидкості вітру  $v_m = 10$  м/сек становить у середньому 97 годин (4 доби) протягом року. Така тривалість імовірних простоїв кранів є істотною, а тому її потрібно враховувати при плануванні перевантажувальних і будівельно-монтажних робіт.

#### 6.4. Апроксимація лінійних залежностей

Задача апроксимації виникає при необхідності описати аналітичним виразом експериментально виявлену залежність між двома чи декількома величинами. Оскільки експериментальні дані завжди мають випадковий розкид, такі залежності є стохастичними залежностями між випадковими величинами. Аналітичний опис експериментальних залежностей доцільно виконувати згідно з принципами методу найменших квадратів за методиками, викладеними в розділі 4.

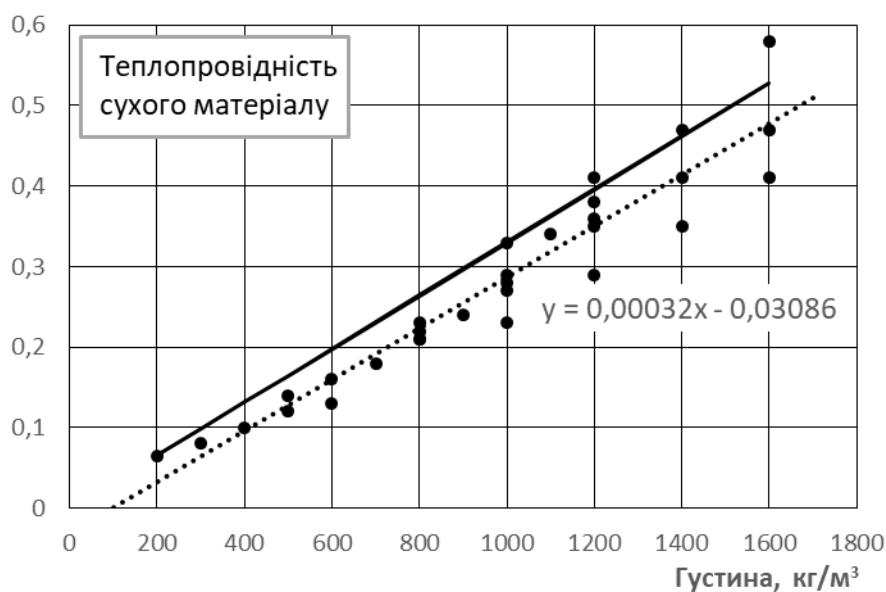
Найпростішою задачею аналітичного опису емпіричних залежностей є апроксимація лінійною функцією. Розглянемо залежність коефіцієнтів теплопровідності від середньої густини легких бетонів. В якості дослідних даних використані наведені в ДСТУ Б В.2.6-189:2013 "Методи вибору теплоізоляційного матеріалу для утеплення будівель" коефіцієнти теплопровідності легких бетонів різного виду з різними значеннями середньої густини. В таблиці ДСТУ, витяг з якої наведено на рисунку 6.6, вказані коефіцієнти теплопровідності в сухому стані, в нормальних та вологих умовах експлуатації для 30 бетонів (ніздрюватий, керамзитовий та шлакобетон) з різними значеннями середньої густини: від 200 кг/м<sup>3</sup> для теплоізоляційного ніздрюватого бетону до 1600 кг/м<sup>3</sup> для шлакобетону.

Вихідні дані занесені в комірки В 5...Е 34 робочого аркуша "6.4" з файлу "Аналіз\_даних.xlsx", копія якого наведена на рисунку 6.6. За цими даними на рисунку 6.7 побудовані графіки залежностей теплопровідності в сухому стані, в нормальних та вологих умовах експлуатації від густини легкого бетону. Розташування дослідних точок на побудованих графіках вказує на наявність тісних лінійних залежностей між проаналізованими величинами, які можна описати рівняннями лінійної регресії виду (4.2). Тіснота лінійного зв'язку підтверджується також наведеними в комірках С 36...Е 36 коефіцієнтами кореляції між теплопровідністю і густиною, дорівнюють  $R = 0,96...0,97$ . Вони обчислені за функцією *KORPEL(...)*.

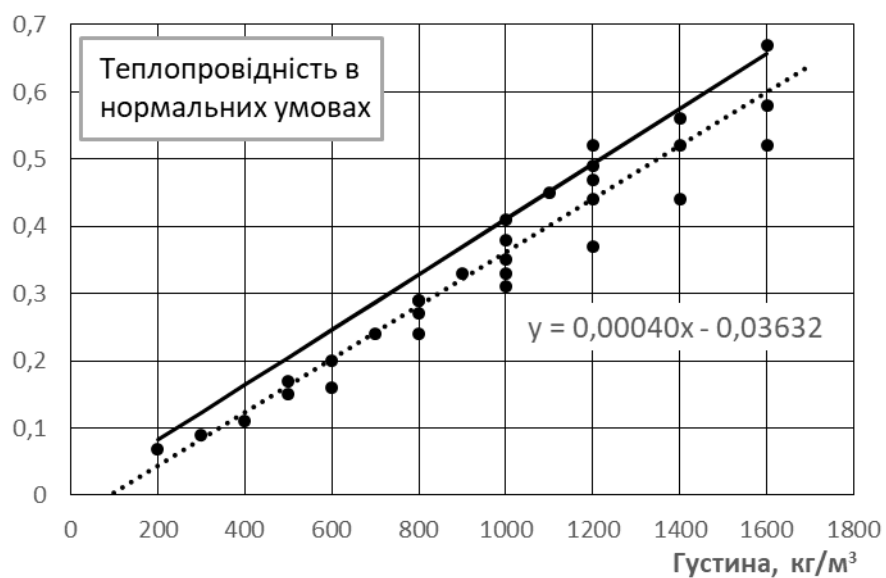
Коефіцієнти  $A$ ,  $B$  рівнянь лінійної регресії для трьох режимів експлуатації бетону обчислені за функціями *ОТРЕЗОК(...)* і *НАКЛОН(...)* та наведені в комірках С 38...Е 39 робочого аркуша з рисунка 6.6. Ті ж самі значення отримані на графіках залежностей теплопровідності від середньої густини, наведених на рисунках 6.7. Для цього була виконана команда "добавить линию тренда" з міткою "показывать уравнение на диаграмме". Використані способи обчислення коефіцієнтів рівняння регресії реалізують метод найменших квадратів та формули (4.3).

	A	B	C	D	E	F	G	H	I	J	K	
1	<b>ТЕПЛОПРОВІДНІСТЬ ЛЕГКИХ БЕТОНІВ</b>											
2												
3	<b>Вид бетону</b>	<b>Густина кг/м3</b>	<b>Умови експлуатації</b>			<b>Умови експлуатації</b>			<b>Умови експлуатації</b>			
4			<b>С</b>	<b>Н</b>	<b>В</b>	<b>С</b>	<b>Н</b>	<b>В</b>	<b>С</b>	<b>Н</b>	<b>В</b>	
5	Нідрюваті бетони	200	0,065	0,069	0,074	0,066	0,084	0,096	1,54	21,74	29,73	
6		300	0,08	0,09	0,10	0,099	0,126	0,144	23,75	40,00	44,00	
7		400	0,10	0,11	0,13	0,132	0,168	0,192	32,00	52,73	47,69	
8		500	0,12	0,15	0,16	0,165	0,210	0,240	37,50	40,00	50,00	
9		600	0,13	0,16	0,18	0,198	0,252	0,288	52,31	57,50	60,00	
10		700	0,18	0,24	0,27	0,231	0,294	0,336	28,33	22,50	24,44	
11		800	0,21	0,27	0,30	0,264	0,336	0,384	25,71	24,44	28,00	
12		900	0,24	0,33	0,36	0,297	0,378	0,432	23,75	14,55	20,00	
13		1000	0,29	0,38	0,44	0,330	0,420	0,480	13,79	10,53	9,09	
14		1100	0,34	0,45	0,51	0,363	0,462	0,528	6,76	2,67	3,53	
15	1200	0,38	0,49	0,55	0,396	0,504	0,576	4,21	2,86	4,73		
16	Керамзитобетони	500	0,14	0,17	0,23	0,165	0,210	0,240	17,86	23,53	4,35	
17		600	0,16	0,20	0,26	0,198	0,252	0,288	23,75	26,00	10,77	
18		800	0,21	0,24	0,31	0,264	0,336	0,384	25,71	40,00	23,87	
19		1000	0,27	0,33	0,41	0,330	0,420	0,480	22,22	27,27	17,07	
20		1200	0,36	0,44	0,52	0,396	0,504	0,576	10,00	14,55	10,77	
21		1400	0,47	0,56	0,65	0,462	0,588	0,672	-1,70	5,00	3,38	
22		1600	0,58	0,67	0,79	0,528	0,672	0,768	-8,97	0,30	-2,78	
23		800	0,23	0,29	0,35	0,264	0,336	0,384	14,78	15,86	9,71	
24		1000	0,33	0,41	0,47	0,330	0,420	0,480	0,00	2,44	2,13	
25		1200	0,41	0,52	0,58	0,396	0,504	0,576	-3,41	-3,08	-0,69	
26	800	0,22	0,29	0,35	0,264	0,336	0,384	20,00	15,86	9,71		
27	1000	0,28	0,35	0,41	0,330	0,420	0,480	17,86	20,00	17,07		
28	Шлакбетони	1000	0,23	0,31	0,37	0,330	0,420	0,480	43,48	35,48	29,73	
29		1200	0,29	0,37	0,44	0,396	0,504	0,576	36,55	36,22	30,91	
30		1400	0,35	0,44	0,52	0,462	0,588	0,672	32,00	33,64	29,23	
31		1600	0,41	0,52	0,63	0,528	0,672	0,768	28,78	29,23	21,90	
32		1200	0,35	0,47	0,52	0,396	0,504	0,576	13,14	7,23	10,77	
33		1400	0,41	0,52	0,58	0,462	0,588	0,672	12,68	13,08	15,86	
34		1600	0,47	0,58	0,64	0,528	0,672	0,768	12,34	15,86	20,00	
35	Коефіцієнти кореляції											
36		0,958	0,966	0,971								
37	Підібрані параметри			Прийняті параметри								
38	A=	-0,03086	-0,03632	-0,03541	0	0	0					
39	B=	0,00032	0,00040	0,00045	0,00033	0,00042	0,00048					
40	<b>Числові характеристики</b>											
41	Min	200,0	0,07	0,07	0,07	0,07	0,08	0,10	-8,97	-3,08	-2,78	
42	Max	1600,0	0,58	0,67	0,79	0,53	0,67	0,77	52,31	57,50	60,00	
43	M=	966,7	0,28	0,35	0,40	0,32	0,41	0,46	18,89	21,60	19,50	
44	S=	385,4	0,13	0,16	0,18	0,13	0,16	0,18	14,44	15,50	15,80	
45	V=	0,40	0,46	0,46	0,45	0,40	0,40	0,40	0,76	0,72	0,81	
46									Довірчі межі відхилень			
47									Dmin	-4,78	-3,82	-6,42
48									Dmax	42,57	47,01	45,42

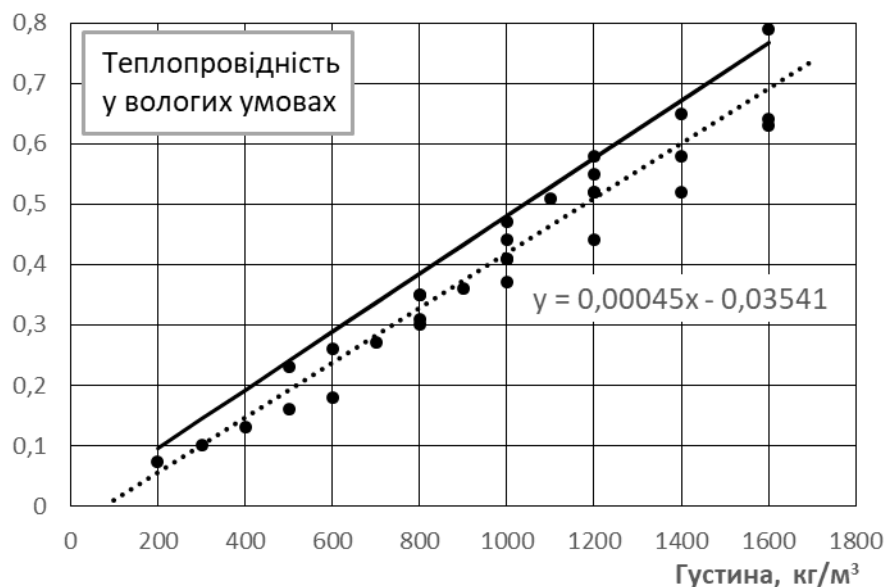
Рис. 6.6. Робочий аркуш Microsoft Excel з прикладом апроксимації експериментальних залежностей лінійними функціями



Розрахункова залежність:  
 $\lambda = 0,00033 \times \rho$   
 $M_{\Delta} = 18,9\%$   
 $\Delta_{\min} = -4,8\%$   
 $\Delta_{\max} = +42,6\%$



Розрахункова залежність:  
 $\lambda = 0,00041 \times \rho$   
 $M_{\Delta} = 18,7\%$   
 $\Delta_{\min} = -6,1\%$   
 $\Delta_{\max} = +43,5\%$



Розрахункова залежність:  
 $\lambda = 0,00048 \times \rho$   
 $M_{\Delta} = 19,5\%$   
 $\Delta_{\min} = -6,4\%$   
 $\Delta_{\max} = +45,4\%$

Рис. 6.7. Залежності теплопровідності від густини легких бетонів

Отримані у результаті розрахунків від'ємні значення коефіцієнтів  $A$  вказують на те, що при наближенні середньої густини бетону до нуля теплопровідність може приймати фізично неможливі від'ємні значення. Другим недоліком формул, отриманих за методом найменших квадратів і відображених на рисунках 6.7 пунктирними лініями, є те, що вони описують досліджені залежності в середньому, тобто у половині випадків дають значення теплопровідності менші від реальних. Це призведе до небажаного зменшення проектного шару теплоізоляції.

З метою встановлення простих залежностей, які забезпечують запас надійності при їх використанні, в комірках F 5...H 34 робочого аркуша реалізовані обчислення за лінійним рівнянням (4.2) при коефіцієнтах, встановлених у комірках F 38...H 39. Отримані залежності зображені на графіках рисунка 6.7 суцільними лініями. З графіків видно, що вони проходять дуже близько до верхньої межі дослідних точок, тобто у переважній більшості випадків завищують значення теплопровідності. Використання цих залежностей призведе до підвищення рівня теплової надійності огорожень унаслідок зростання товщини утеплювача.

Для контролю якості апроксимації в комірках I 5...H 34 за формулою (4.7) обчислені відносні відхилення  $\Delta_B$  апроксимуючих ліній з комірок F 5...H 34 від дослідних даних з комірок C 5...E 34. В рядках 41...45 за відповідними формулами з підрозділу 3.3 та функціями Microsoft Excel, що їх реалізують, визначені числові характеристики усіх проаналізованих вибірок. В комірках I 47...K 48 за формулами (4.8) обчислені довірчі межі відхилень апроксимуючих функцій від дослідних даних для довірчої імовірності  $P = 0,9$ , тобто при  $t = 1,64$ .

Коефіцієнти апроксимуючих прямих  $A$ ,  $B$  в комірках F 38...H 39 підібрані вручну з точністю до двох знаків з урахуванням положення цих ліній на графіках рисунка 6.7, та довірчих меж отриманих відхилень з комірок I 47...K 48. Враховуючи близькість апроксимуючих прямих, підібраних методом найменших квадратів (пунктири на графіках рисунка 6.7), до початку координат, при підборі коефіцієнтів для суцільних ліній і комірках F 38...H 38 встановлено  $A = 0$ . Тому рівняння прямих, підібраних в запас надійності, мають вигляд

$$\lambda = B \times \rho. \quad (6.3)$$

Такі рівняння з конкретними значеннями коефіцієнта  $B$  наведені справа від відповідних графіків на рисунку 6.7.

Аналіз продубльованих на тому ж рисунку параметрів відхилень апроксимуючих прямих (6.3) від дослідних даних вказує на досить високу точність і цілком достатні запаси апроксимації. Середній запас змінюється від 18,7% до 19,5%. Наведені на рисунках 6.6 і 6.7 90% довірчі межі відхилень змінюються від 4,8% до 6,4% у бік заниження та від 42,6% до 45,4% у бік завищення значень теплопровідності (6.3) порівняно з експериментальними даними. Прийнята двостороння довірча імовірність 0,9 означає, що заниження значень теплопровідності на 4,8...6,4% може реалізуватися не більше ніж у 5% випадків визначення теплопровідності за формулами виду (6.3), наведеними на рисунку 6.7.

Таким чином, розділення за умовами експлуатації дозволило підібрати для залежностей теплопровідності легких бетонів від їх середньої густини достатньо точні й прості апроксимуючі формули у вигляді (6.3). Виконаний приклад може слугувати зразком аналізу та апроксимації різних залежностей лінійними функціями. За викладеною методикою можна підібрати також функцію, графік якої проходить поблизу нижньої межі дослідних точок, якщо саме таке положення створює необхідний запас надійності.

### **6.5. Апроксимація нелінійної залежності типовими функціями**

Методику опису нелінійних залежностей функціями, наявними в Microsoft Excel, пояснимо на прикладі залежності розрахункового опору ніздрюватого бетону від його середньої густини. Міцність ніздрюватого бетону в основному залежить від його пористості (обумовлює частку твердої складової та густину бетону), використаного цементу та інших матеріалів (обумовлюють міцність твердої складової), а також технології твердіння (повітряне чи автоклавне). Характеристики ніздрюватого бетону різних марок визначені за ДСТУ Б В.2.7-45:2010 "Бетони ніздрюваті. Загальні технічні умови". Розрахункові опори при стиску для ніздрюватих бетонів марок від D 300 до D 1200 за середньою густиною вказані в комірках С 8...D 47 робочого аркуша "6.5" з файлу Microsoft Excel "Аналіз\_даних.xlsx", зображення якого наведено на рисунку 6.8. Зі стовпців А, В видно, що в наведених даних враховані мінімально можливі та максимально можливі значення розрахункового опору для бетонів автоклавного і неавтоклавного (повітряного) твердіння. Це обумовило великий розкид дослідних точок, відображений на побудованих за цими даними графіками 6.9 і 6.10.

	A	B	C	D	E	F	G	H		
1	<b>РОЗРАХУНКОВИЙ ОПІР НІЗДРЮВАТИХ БЕТОНІВ</b>									
2										
3	Вид	Зна-	Марка	Опір при	Поліном		Експонента			
4	бетону	чення	за	стиску	функція	Δ%	функція	Δ%		
5			0		0,00		0,13			
6			100		-0,04		0,18			
7			200		0,00		0,25			
8	автоклавний	мінімальні	300	0,45	0,12	73,33	0,34	24,55		
9			400	0,63	0,32	49,21	0,47	25,78		
10			500	0,95	0,60	36,84	0,64	32,22		
11			600	1,30	0,96	26,15	0,89	31,79		
12			700	1,60	1,40	12,50	1,22	23,68		
13			800	1,60	1,92	-20,00	1,68	-5,10		
14			900	3,10	2,52	18,71	2,32	25,29		
15			1000	4,60	3,20	30,43	3,19	30,67		
16			1100	6,00	3,96	34,00	4,39	26,80		
17			1200	7,00	4,80	31,43	6,05	13,60		
18			максимальні	300	0,63	0,12	80,95	0,34	46,11	
19				400	0,95	0,32	66,32	0,47	50,78	
20		500		1,60	0,60	62,50	0,64	59,76		
21		600		1,60	0,96	40,00	0,89	44,58		
22		700		3,10	1,40	54,84	1,22	60,61		
23		800		3,10	1,92	38,06	1,68	45,75		
24		900		4,60	2,52	45,22	2,32	49,66		
25		1000		6,00	3,20	46,67	3,19	46,85		
26		1100		7,70	3,96	48,57	4,39	42,96		
27		1200		7,70	4,80	37,66	6,05	21,45		
28		неавтоклавний		мінімальні	300	0,30	0,12	60,00	0,34	-13,17
29					400	0,63	0,32	49,21	0,47	25,78
30			500		0,95	0,60	36,84	0,64	32,22	
31			600		0,95	0,96	-1,05	0,89	6,66	
32	700		1,30		1,40	-7,69	1,22	6,07		
33	800		1,30		1,92	-47,69	1,68	-29,36		
34	900		2,20		2,52	-14,55	2,32	-5,27		
35	1000		3,10		3,20	-3,23	3,19	-2,88		
36	1100		4,60		3,96	13,91	4,39	4,52		
37	1200		6,00		4,80	20,00	6,05	-0,81		
38	максимальні		300		0,45	0,12	73,33	0,34	24,55	
39			400		0,95	0,32	66,32	0,47	50,78	
40			500	1,30	0,60	53,85	0,64	50,47		
41			600	1,30	0,96	26,15	0,89	31,79		
42			700	1,60	1,40	12,50	1,22	23,68		
43			800	2,20	1,92	12,73	1,68	23,56		
44			900	3,10	2,52	18,71	2,32	25,29		
45			1000	4,60	3,20	30,43	3,19	30,67		
46			1100	7,00	3,96	43,43	4,39	37,26		
47			1200	7,00	4,80	31,43	6,05	13,60		
48			<b>Параметри функцій</b>							
49						A=	0,000004		0,13	
50					B=	-0,0008		0,0032		
51	<b>Характеристики відхилень</b>									
52				M=		32,2		25,8		
53				S=		27,6		20,6		
54				Δmin		-13,0		-8,0		
55				Δmax		77,4		59,6		

Рис. 6.8. Апроксимація залежності розрахункового опору від густини ніздрюватого бетону

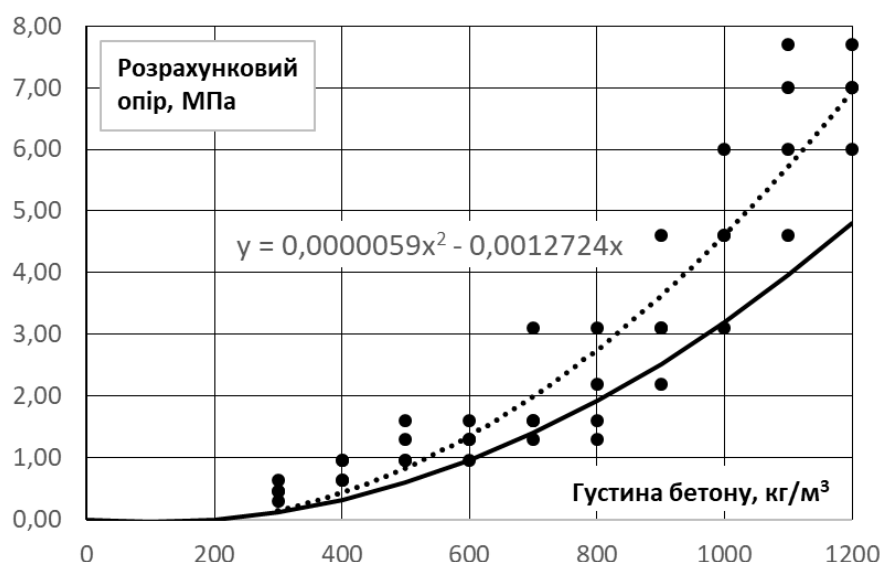


Рис. 6.9. Апроксимація залежності розрахункового опору від густини ніздрюватого бетону квадратичною функцією

Загалом графік залежності розрахункового опору від густини має зростаючий характер з увігнутістю вниз. Спроби описати цю залежність вбудованими в Microsoft Excel нелінійними функціями з використанням команди "добавить линию тренда" та використати ці функції в якості можливих варіантів апроксимації показали наступне:

- лінійна  $Y = A + B \times X$  та логарифмічна залежність виду  $Y = A + B \ln(X)$  явно не відповідають характеру наявного графіка;
- ступенева залежність виду  $Y = A X^B$  дає значення параметра  $B \approx 1$ , тобто по суті вироджується в квадратичну залежність, і при цьому незадовільно відображає характер дослідного графіка;
- квадратна парабола  $Y = A \times X^2 + B \times X$ , яка проходить через початок координат, має вигляд, близький до характеру зміни дослідного графіка, і може розглядатися в якості апроксимуючої функції;
- експоненціальна залежність виду  $Y = A e^{B \times X}$  за характером близька до графіка зміни розрахункового опору і також може розглядатися в якості апроксимуючої функції.

Викладені результати попереднього аналізу графіка з рисунка 6.9 вказують на доцільність розгляду двох апроксимуючих функцій:

$$\text{квадратної параболи} \quad Y = A \times X^2 + B \times X \quad (6.4)$$

$$\text{та експоненти} \quad Y = A e^{B \times X}. \quad (6.5)$$

Результати визначення параметрів цих функцій з використанням команди "добавить линию тренда" наведені на рисунках 6.9 і 6.10. Підібрані функції зображені на графіках пунктирними лініями, а їх параметри вказані у наведених рівняннях. Звичайно, функції підібрані за методом найменших квадратів, описують досліджену залежність "у середньому". Для впевненого прогнозування в запас надійності необхідно провести апроксимуючі функції поблизу нижньої межі дослідних точок, щоб вони давали мінімально можливі значення розрахункового опору. Ця задача вирішена шляхом підбору параметрів, як це було зроблено в задачі підбору лінійної функції з попереднього підрозділу.

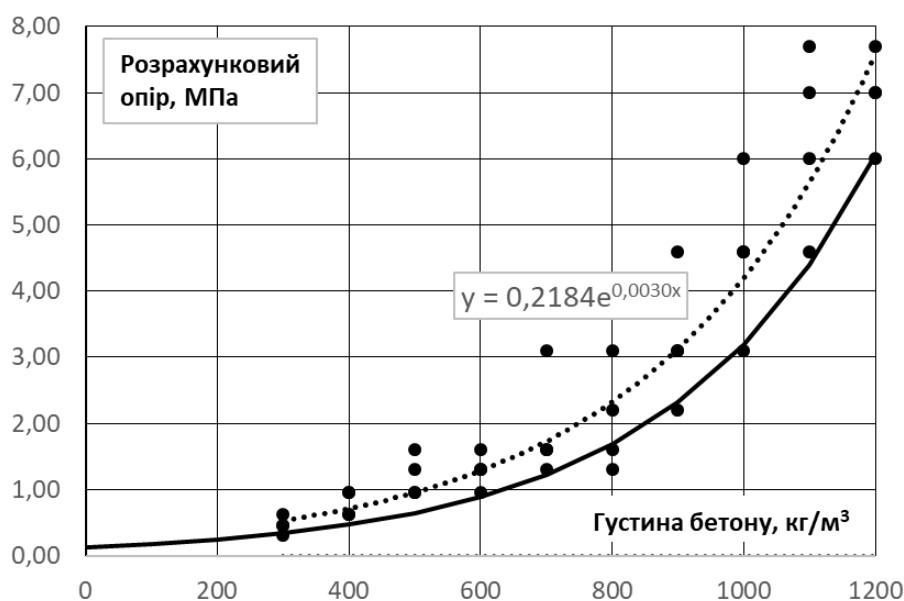


Рис. 6.10. Апроксимація залежності розрахункового опору від густини ніздрюватого бетону експоненціальною функцією

Якість вибору параметрів функцій (6.4) і (6.5) контролювалася візуально та за характеристиками точності, обчисленими в комірках F 52...H 56 робочого аркуша Excel з рисунка 6.8. Для цього в комірках E 5...E 47 і G 5...G 47 за формулами (6.4) і (6.5) обчислені теоретичні значення функцій з параметрами з комірок E 49...E 50. Комірки F 8...F 47 і H 8...H 47 містять вибірки відносних відхилень дослідних значень розрахункового опору бетону від теоретичних, а в комірках F 52...F 55 і H 52...H 55 визначені числові характеристики цих відхилень. Зміна знаку відносних відхилень забезпечила їх логічне сприйняття: знак "плюс" вказує на запас надійності (теоретичні значення є меншими за дослідні), а знак "мінус" символізує небажану подію, коли апроксимуюча функція завищує значення розрахункового опору бетону.

У результаті виконаного підбору параметрів квадратична функція (6.4) отримала вигляд, зображений на рисунку 6.9 суцільною лінією:

$$R_b = 0,000004 \times \rho^2 - 0,0008 \times \rho = 4 \times 10^{-6} \rho \times (\rho - 200) . \quad (6.6)$$

Отримана функція досить проста, але вона має два недоліки. По-перше, функція (6.6) перетинає ось абсцис при значенні густини бетону  $\rho = 200 \text{ кг/м}^3$ , а при менших  $\rho$  приймає від'ємні значення, що видно з комірки Е 6 таблиці з рисунка 6.8. Другий недолік полягає в надмірному заниженні розрахункового опору при великих значеннях густини бетону. Спроба надати функції (6.4) більшої кривизни призводить до подальшого зростання зони від'ємних значень розрахункового опору при малих значеннях густини бетону.

Підбір параметрів експоненціальної залежності (6.5) завершився отриманням виразу

$$R_b = 0,13 e^{0,032\rho} , \quad (6.7)$$

графік якого зображений суцільною лінією на рисунку 6.10.

Навіть чисто візуально експоненціальна функція (6.7) з рисунка 6.10 краще описує дослідну залежність, ніж квадратична функція (6.6) з рисунка 6.8. На користь експоненти (6.7) свідчить також її адекватна асимптотична поведінка при  $\rho \rightarrow 0$ , менший довірчий інтервал відхилень від дослідних точок (68% замість 90% для функції (6.6)) та одночасно – менші відхилення не в запас надійності ((-8% замість -13% для функції (6.6)). Виходячи з викладеного, для опису залежності розрахункового опору ніздрюватих бетонів від їх середньої густини в запас надійності слід використовувати експоненціальну функцію (6.7).

## 6.6. Апроксимація двофакторної залежності

За результатами експериментального проектування 42 прокатних балок із сталі С 245 встановлена таблична залежність маси прокатних балок від прольоту та навантаження на них. Балки прольотами  $L = 3 \dots 24 \text{ м}$  виконані з прокатних двотаврів балкового типу і сприймають погонні навантаженнями  $q = 5 \dots 150 \text{ кН/м}$ . З умов міцності та жорсткості підібрані перерізи балок від 10Б1 до 100Б4, що охоплює увесь сортамент прокатних двотаврів балкового типу. Для деяких співвідношень прольотів і навантажень переріз балки підібрати неможливо, оскільки несуча здатність найбільшого в сортаменті двотавра 100Б4 є недостатньою.

Значення погонної маси запроєктованих балок в кг/м та занесені на робочий аркуш "6.6" з файлу Microsoft Excel "Аналіз\_даних.xlsx", фрагмент якого наведений на рисунку 6.11. Дані комірок В17...І23 по суті є табличною залежністю маси від прольоту та погонного навантаження на балки. Для зручності користування отриману таблицю слід описати аналітичною залежністю, яка дозволять при проектуванні прогнозувати погонну масу балок через проліт і погонне навантаження.

	A	B	C	D	E	F	G	H	I
13	<b>ПОГОННА МАСА ПРОКАТНИХ БАЛОК</b>								
14									
15	Погонна маса (дослідна) в кг/м								
16		<b>3</b>	<b>6</b>	<b>9</b>	<b>12</b>	<b>15</b>	<b>18</b>	<b>21</b>	<b>24</b>
17	5	8,1	15,4	28,0	43,3	59,8	80,7	106,2	129,3
18	25	15,4	36,6	59,8	89,0	129,3	159,5	213,8	258,2
19	50	25,8	54,7	97,9	144,2	177,9	230,6	314,5	
20	75	31,2	73,0	129,3	177,9	230,6	314,5		
21	100	36,6	89,0	144,2	213,8	285,7			
22	125	38,9	106,2	177,9	258,2				
23	150	48,1	115,6	194,0	285,7				
24									
25		A	B	C				Цільова	
26		1,1502	0,0457	-1,885				2761,1	
27									
28	Погонна маса (теоретична) в кг/м								
29		<b>3</b>	<b>6</b>	<b>9</b>	<b>12</b>	<b>15</b>	<b>18</b>	<b>21</b>	<b>24</b>
30	5	6,8	17,2	29,5	43,7	59,7	77,5	97,2	118,8
31	25	17,4	40,9	68,4	100,1	135,8	175,7	219,7	267,9
32	50	25,4	58,6	97,5	142,3	192,9	249,3	311,6	
33	75	31,6	72,1	119,9	174,7	236,7	305,8		
34	100	36,7	83,6	138,7	202,0	273,6			
35	125	41,3	93,7	155,3	226,1				
36	150	45,4	102,8	170,3	247,8				
37									
38	Відхилення у відсотках								
39		<b>3</b>	<b>6</b>	<b>9</b>	<b>12</b>	<b>15</b>	<b>18</b>	<b>21</b>	<b>24</b>
40	5	-16,7	11,9	5,5	0,9	-0,2	-3,9	-8,4	-8,2
41	25	13,2	11,6	14,4	12,4	5,1	10,2	2,8	3,7
42	50	-1,5	7,1	-0,4	-1,3	8,4	8,1	-0,9	
43	75	1,2	-1,2	-7,3	-1,8	2,6	-2,8		
44	100	0,4	-6,1	-3,8	-5,5	-4,2			
45	125	6,2	-11,8	-12,7	-12,4				
46	150	-5,6	-11,1	-12,2	-13,3				

Рис. 6.11. Апроксимація залежності погонної маси балок від прольоту та навантаження

Поставлене завдання розв'язане за методикою, викладеною в підрозділі 4.3. З метою обґрунтування виду залежностей погонної маси балок від обох визначальних аргументів – прольоту  $L$  та навантаження  $q$  на рисунку 6.12 за даними комірок А 16...І 23 показаного вище робочого аркуша Excel зображені роздільні залежності погонної маси від цих величин. Залежності від прольоту збудовані для фіксованих навантажень 5, 50 і 150 кН/м, а залежності від навантаження – для прольотів 3, 9 і 15 м.

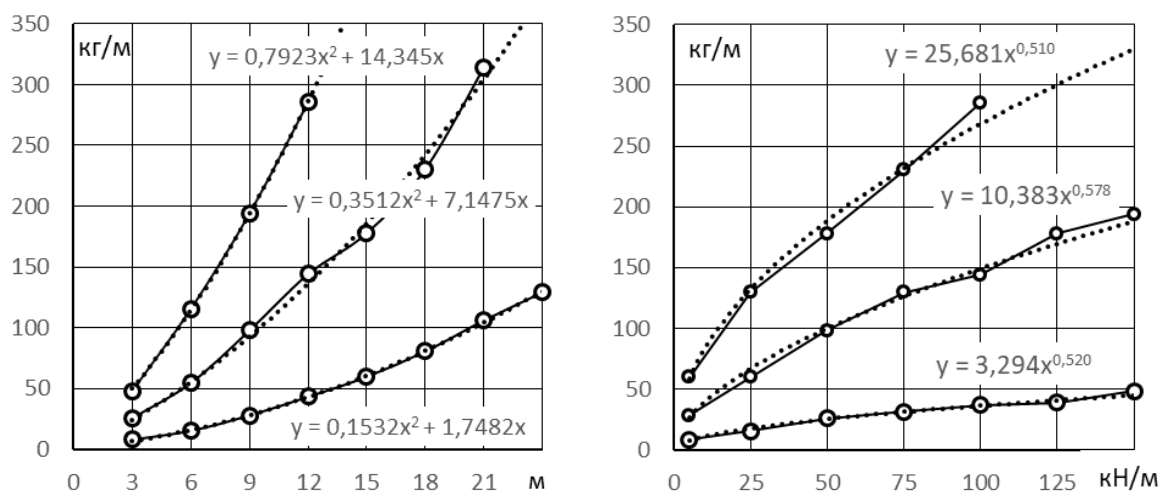


Рис. 6.12. Залежності погонної маси балок від прольоту та навантаження

З рисунка видно, що залежності погонної маси від прольоту балок мають характер квадратних парабол, які проходять через початок координат і добре описуються апроксимуючими лініями тренду, рівняння яких наведені на графіках:

$$P = A \times L + B \times L^2 . \quad (6.8)$$

Такий характер залежностей (6.8) обумовлений квадратичною залежністю максимального згинального моменту від прольоту балки. Відхилення експериментальних даних від ліній тренду викликані значним кроком зміни геометричних характеристик поперечних перерізів у сортаменті, що й призводить до округлень при підборі перерізів балок.

Наведені на рисунку 6.12 графіки залежностей погонної маси балок від навантаження мають незначну випуклість угору. З графіків видно, що ці залежності задовільно описуються ступеневими функціями виду:

$$p = \beta \times q^\alpha , \quad (6.9)$$

у яких параметр  $\beta$  задає вертикальний масштаб, а показник ступеня змінюється в межах  $0,51 \leq \alpha \leq 0,58$ . Близькість параметра  $\alpha$  до величини 0,5 дозволяє замінити функцію (6.9) функцією з квадратним коренем

$$p = \beta \times \sqrt{q} . \quad (6.10)$$

Вид об'єднаної залежності погонної маси  $p$  від прольоту  $L$  й навантаження на балку  $q$  можна отримати, перемноживши (6.8) на (6.10) і додавши коригуючий параметр  $C$ :

$$p = (A \times L + B \times L^2) \times \sqrt{q} + C . \quad (6.11)$$

Параметри  $A$ ,  $B$ ,  $C$  залежності (6.11) обчислені на основі методу найменших квадратів [2] з використанням функції Microsoft Excel "Поиск решения" [5]. Хід обчислень показано на копії робочого аркуша, наведеній на рисунку 6.11.

Перший блок комірок А 16...І 23 містить отримані в результаті експериментального проектування значення погонної маси балок для різних прольотів і навантажень. Другий блок розрахункової таблиці в комірках А 29...І 36 містить значення погонної маси балок, обчислені за формулою (6.11) при значеннях параметрів  $A$ ,  $B$ ,  $C$  з комірок В 26...D 26. Третій блок розрахункової таблиці в комірках А 39...І 46 містить відсотки відхилень теоретичної погонної маси балок з комірок А 29...І 36 від фактично отриманих значень з комірок А 16...І 23.

Цільова функція для мінімізації суми квадратів обчислена в комірці Н 26 з використанням функції СУММКВ(...), як сума квадратів відсотків відхилень теоретичних значень погонної маси балок від фактичних, тобто сума квадратів даних комірок В 40...І 46.

Постановка задачі "поиск решения" має таке формулювання: знайти мінімально можливе значення цільової функції в комірці Н 26, змінюючи значення параметрів  $A$ ,  $B$ ,  $C$  в комірках В 26...D 26. Excel розв'язує цю задачу в автоматичному режимі за вбудованим алгоритмом мінімізації функцій. В результаті розв'язання для даного випадку розрахунку отримані параметри  $A = 1,1502$ ;  $B = 0,0457$ ;  $C = -1,885$  та значення цільової функції 2761,1. Теоретичні значення погонної маси балок за (6.11) та відсотки їх відхилень від фактичних даних, що відповідають обчисленим параметрам  $A$ ,  $B$ ,  $C$ , наведені в розрахунковій таблиці Excel на рисунку 6.11.

З таблиці видно, що відсотки відхилень змінюються від  $-16,7\%$  до  $+14,4\%$ , а їх середнє значення досить близьке до нуля. Отримані відхилення для різних комбінацій прольоту й навантаження на балку змінюються хаотично, що обумовлено розкидом фактичних даних у результаті сортаментних округлень при підборі перерізів балок.

З метою спрощення отриманої формули (6.11) виконане округлення її параметрів. Для цього в комірки B 26...D 26 розрахункової таблиці вводилися округлені значення параметрів  $A$ ,  $B$ ,  $C$ , і відслідковувалися зміни цільової функції в комірці H 26 та відсотків відхилень теоретичних значень маси балок від дослідних у комірках B 40...I 46. Зміни результатів апроксимації при округленні параметрів представлені на рисунку 6.13.

	A	B	C	D	E	F	G	H	I
25		A	B	C				Цільова	
26		1,1500	0,0460	-1,900				2762,1	
27									
28	Погонна маса (теоретична) в кг/м								
29		<b>3</b>	<b>6</b>	<b>9</b>	<b>12</b>	<b>15</b>	<b>18</b>	<b>21</b>	<b>24</b>
30	5	6,7	17,2	29,6	43,8	59,8	77,7	97,5	119,1
31	25	17,4	40,9	68,5	100,2	136,1	176,1	220,3	268,6
32	50	25,4	58,6	97,6	142,5	193,3	249,9	312,3	
33	75	31,6	72,2	120,0	175,0	237,1	306,4		
34	100	36,7	83,7	138,9	202,3	274,1			
35	125	41,3	93,8	155,5	226,4				
36	150	45,4	102,9	170,5	248,2				
37									
38	Відхилення у відсотках								
39		<b>3</b>	<b>6</b>	<b>9</b>	<b>12</b>	<b>15</b>	<b>18</b>	<b>21</b>	<b>24</b>
40	5	-16,8	11,9	5,6	1,1	0,0	-3,7	-8,2	-7,9
41	25	13,1	11,7	14,5	12,6	5,3	10,4	3,0	4,0
42	50	-1,5	7,1	-0,3	-1,2	8,6	8,4	-0,7	
43	75	1,2	-1,1	-7,2	-1,6	2,8	-2,6		
44	100	0,4	-6,0	-3,7	-5,4	-4,1			
45	125	6,2	-11,7	-12,6	-12,3				
46	150	-5,6	-11,0	-12,1	-13,1				

Рис. 6.13. Результати апроксимації при округлених значеннях параметрів

Порівняння розрахункових таблиць з рисунків 6.11 і 6.13 показує, що виконані округлення параметрів не внесли помітних змін у результати апроксимації. Значення цільової функції збільшилося з 2761,1 до 2762,1, а межі зміни відхилень апроксимуючої функції від дослідних даних зросли з  $-16,7...+14,4\%$  до  $-16,8...+14,5\%$ . Тому для опису залежності погонної маси прокатних балок  $p$  (кг/м) від їх прольоту  $L$  (м) та погонного навантаження  $q$  (кН/м) рекомендується використовувати формулу:

$$p = (1,15 \times L + 0,046 \times L^2) \times \sqrt{q} - 1,9 . \quad (6.12)$$

Цільова функція D, рівна сумі квадратів відхилень теоретичних значень від дослідних, дозволяє обчислити імовірну похибку апроксимації:

$$\Delta = 1,64 \times \sqrt{D/N} = 1,64 \times \sqrt{2761,1/42} = 13,3\% . \quad (6.13)$$

Значення 1,64 відповідає довірчій імовірності 0,9, а кількість даних  $N = 42$ .

Отриманий розкид відхилень відповідає кроку зміни геометричних характеристик поперечних перерізів у сортаменті балкових двотаврів. Він обумовлений не похибками апроксимації, а округленнями характеристик до сортаментних значень при підборі перерізів балок.

### 6.7. Порівняння теоретичних результатів з дослідними даними

В якості прикладу верифікації теорії за даними експерименту розглянемо залежність теплопровідності від середньої густини матеріалів. Зменшити теплопровідність матеріалу можна за рахунок збільшення кількості пор, що одночасно зменшує також і середню густину. З курсу будівельного матеріалознавства відома формула В.П. Некрасова, яка встановлює залежність теплопровідності від середньої густини кам'яних матеріалів. Якщо виражати теплопровідність  $\lambda$  у Вт/(м×К), а густину бетону  $\rho$  в кг/м<sup>3</sup>, ця формула приймає вигляд:

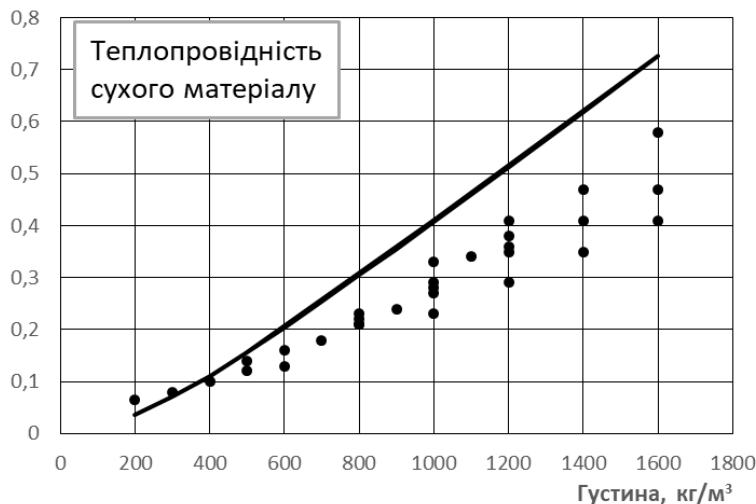
$$\lambda = 1,16 \sqrt{(0,0196 + 0,00000022 \rho^2)} - 0,16. \quad (6.14)$$

Значення теплопровідності (6.14) будемо вважати теоретичними. Експериментальними значеннями вважаємо дані з ДСТУ Б В.2.6-189:2013 "Методи вибору теплоізоляційного матеріалу для утеплення будівель", де наведені коефіцієнти теплопровідності 30 легких бетонів різного виду (ніздрюватий, керамзитовий та шлакобетон) зі значеннями середньої густини: від 200 кг/м<sup>3</sup> 1600 кг/м<sup>3</sup>. Таблиця даних сформована в комітках В 5...F 34 робочого аркуша "6.7" файлу "Аналіз\_даних.xlsx", фрагмент якого зображений на рисунку 6.14. Ці ж дані були використані в підрозділі 6.4 для виконання прикладу апроксимації лінійною функцією. Додатково до значень коефіцієнтів теплопровідності бетонів в сухому стані, в нормальних та вологих умовах експлуатації в стовпчику С наведені теоретичні значення теплопровідності, обчислені за формулою (6.14) для відповідних значень середньої густини зі стовпчика В.

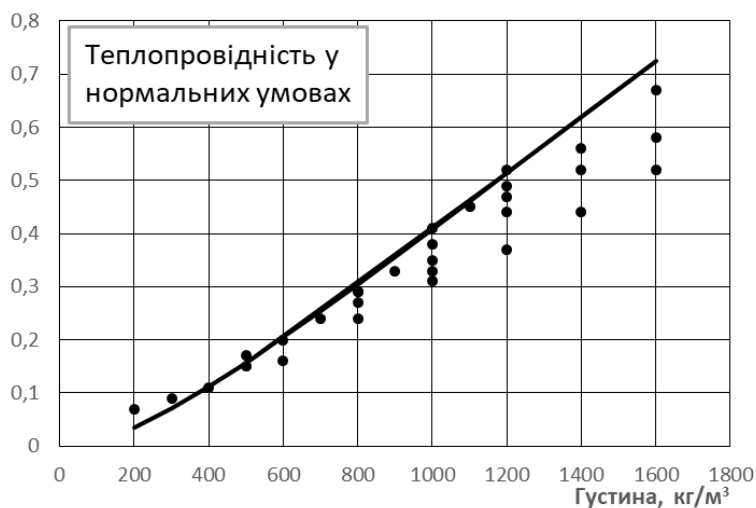
	A	B	C	D	E	F	G	H	I
1	<b>ТЕПЛОПРОВІДНІСТЬ ЛЕГКИХ БЕТОНІВ</b>								
2									
3	<b>Вид</b>	<b>Густин</b>	<b>Теплопровідності в умовах</b>			<b>Відносні похибки, %</b>			
4	<b>бетону</b>	<b>a</b>	<b>теорет</b>	<b>C</b>	<b>H</b>	<b>B</b>	<b>C</b>	<b>H</b>	<b>B</b>
5	<b>Ніздрюваті бетони</b>	200	0,035	0,065	0,069	0,074	-45,4	-48,6	-52,0
6		300	0,070	0,08	0,09	0,10	-12,2	-21,9	-29,7
7		400	0,112	0,10	0,11	0,13	11,5	1,4	-14,2
8		500	0,157	0,12	0,15	0,16	30,7	4,6	-2,0
9		600	0,205	0,13	0,16	0,18	57,4	27,9	13,7
10		700	0,254	0,18	0,24	0,27	41,1	5,9	-5,9
11		800	0,305	0,21	0,27	0,30	45,0	12,8	1,5
12		900	0,356	0,24	0,33	0,36	48,3	7,9	-1,1
13		1000	0,408	0,29	0,38	0,44	40,6	7,3	-7,3
14		1100	0,460	0,34	0,45	0,51	35,3	2,3	-9,8
15	1200	0,513	0,38	0,49	0,55	34,9	4,7	-6,8	
16	<b>Керамзитобетони</b>	500	0,157	0,14	0,17	0,23	12,0	-7,7	-31,8
17		600	0,205	0,16	0,20	0,26	27,9	2,3	-21,3
18		800	0,305	0,21	0,24	0,31	45,0	26,9	-1,7
19		1000	0,408	0,27	0,33	0,41	51,0	23,6	-0,5
20		1200	0,513	0,36	0,44	0,52	42,4	16,5	-1,4
21		1400	0,619	0,47	0,56	0,65	31,7	10,5	-4,8
22		1600	0,726	0,58	0,67	0,79	25,1	8,3	-8,2
23		800	0,305	0,23	0,29	0,35	32,4	5,0	-13,0
24		1000	0,408	0,33	0,41	0,47	23,6	-0,5	-13,2
25		1200	0,513	0,41	0,52	0,58	25,1	-1,4	-11,6
26	800	0,305	0,22	0,29	0,35	38,4	5,0	-13,0	
27	1000	0,408	0,28	0,35	0,41	45,6	16,5	-0,5	
28	<b>Шлакобетони</b>	1000	0,408	0,23	0,31	0,37	77,3	31,6	10,2
29		1200	0,513	0,29	0,37	0,44	76,8	38,6	16,5
30		1400	0,619	0,35	0,44	0,52	76,8	40,6	19,0
31		1600	0,726	0,41	0,52	0,63	77,0	39,5	15,2
32		1200	0,513	0,35	0,47	0,52	46,5	9,1	-1,4
33		1400	0,619	0,41	0,52	0,58	50,9	19,0	6,7
34	1600	0,726	0,47	0,58	0,64	54,4	25,1	13,4	
35									
36			R=	0,959	0,966	0,970			
37						<b>Числові характеристики</b>			
38						N=	30	30	30
39						Min	-45,4	-48,6	-52,0
40						Max	77,3	40,6	19,0
41						M=	38,3	10,4	-5,2
42						S=	25,6	18,2	15,3
43						V=	0,67	1,75	-2,97
44						Dmin	-3,8	-19,4	-30,3
45						Dmax	80,3	40,3	20,0

Рис. 6.14. Порівняння теоретичних результатів (6.14) з дослідними даними по теплопровідності легких бетонів

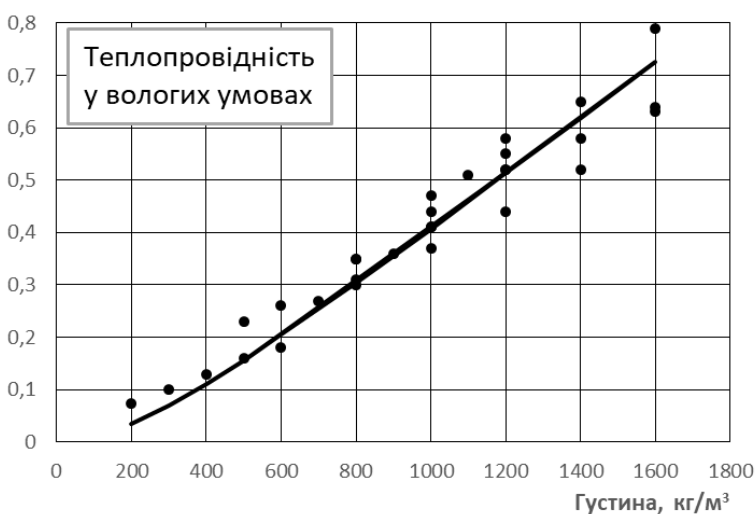
За даними таблиці з розрахункового бланка (рис. 6.14) побудовані залежності теплопровідності від середньої густини, зображені на рисунку 6.15. Крапки відображають дослідні дані, а суцільні лінії – теоретичні значення за формулою Некрасова (6.14).



Для бетонів у сухому стані коефіцієнт теплопровідності змінюється від 0,065 до 0,58



Для бетонів у нормальних умовах експлуатації коефіцієнт теплопровідності змінюється від 0,069 до 0,67



Для бетонів у вологих умовах експлуатації коефіцієнт теплопровідності змінюється від 0,074 до 0,79

Рис. 6.15. Залежності теплопровідності від густини легких бетонів

З ростом вологості теплопровідність легких бетонів також зростає. Наприклад, при густині  $\rho = 1400 \text{ кг/м}^3$  теплопровідність змінюється від  $0,4 \text{ Вт/(м}\times\text{К)}$  для сухого бетону до  $0,6 \text{ Вт/(м}\times\text{К)}$  для бетону, що експлуатується у вологих умовах. Теоретичні значення теплопровідності, обчислені за формулою (6.14) однакові на усіх трьох графіках, оскільки ця формула не враховує вологості матеріалу. З рисунка 6.15 видно, що формула (6.14) істотно завищує теплопровідність сухого матеріалу, задовільно описує верхню межу дослідних точок для теплопровідності бетону в нормальних умовах експлуатації та відображає дослідні точки в середньому для вологих умов експлуатації.

Характер зміни відхилень теоретичних значень теплопровідності від експериментальних можна проаналізувати на графіках з описаного вище рисунка 6.15 та з рисунка 6.16, де наведені залежності теоретичних значень теплопровідності від експериментальних. В обох випадках чітко видно, що при збільшенні коефіцієнта теплопровідності абсолютні різниці між дослідними й теоретичними значеннями зростають. Тому відповідно до рекомендацій підрозділу 4.4, аналізуються відносні відхилення  $\Delta_B$ , обчислені за формулою (4.7) і наведені в трьох останніх стовпчиках розрахункової таблиці, зображеної на рисунку 6.14.

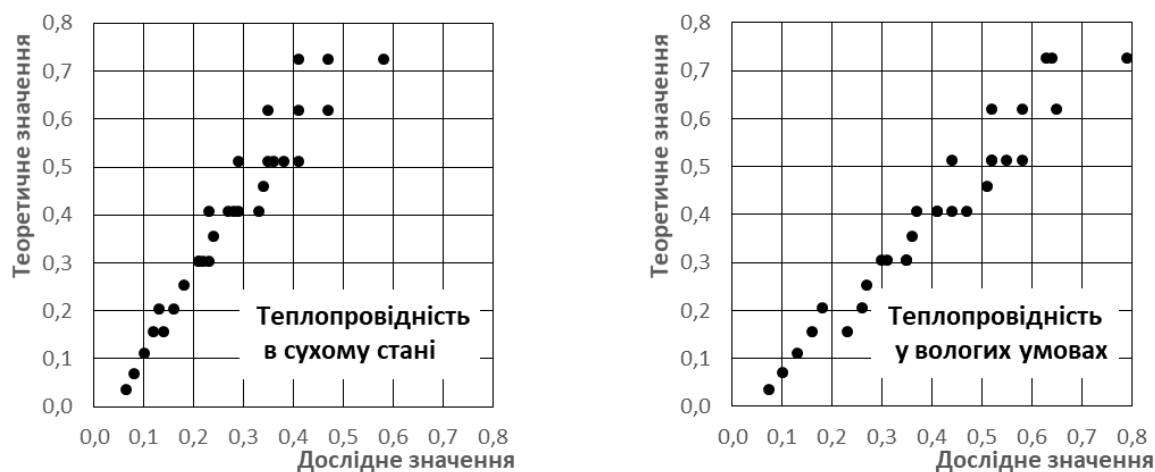


Рис. 6.16. Залежності теоретичних значень теплопровідності від дослідних

Статистична обробка отриманих вибірок відносних відхилень теоретичних значень теплопровідності легких бетонів від дослідних виконана в середовищі Microsoft Excel за методикою, описаною в розділі 3 та використаною у попередніх прикладах статистичної обробки випадкових величин. Отримані в результаті обробки статистичні характеристики відносних відхилень (розмахи вибірок, середні значення й стандарти) наведені в комірках G 38...I 43 розрахункової таблиці з рисунка 6.14. Слід

звернути увагу, що при близьких до нуля середніх значеннях відхилень їх коефіцієнти варіації приймають дуже великі і навіть від'ємні значення, тому їх використання є взагалі недоцільним.

На рисунку 6.17 наведені гістограми розподілу відносних відхилень теоретичних значень теплопровідності бетонів від дослідних. Вони збудовані лише для виявлення форми розподілів, а тому вздовж осі ординат відкладена не густина розподілу, а кількості спостережених випадків у кожному з інтервалів. Загальний вигляд гістограм показує, що їх можна орієнтовно описати нормальним законом розподілу, що дозволяє скористатися формулами (4.8) для визначення довірчих меж відхилень теоретичних значень від експериментальних.

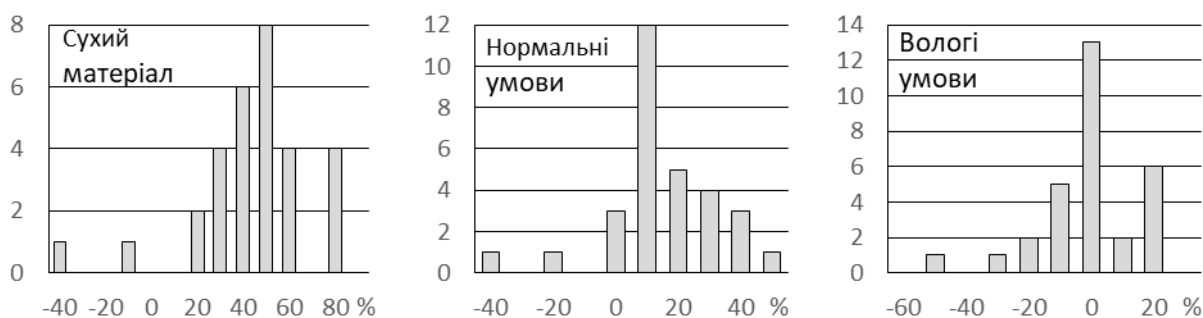


Рис. 6.17. Гістограми розподілу відносних відхилень теоретичних значень теплопровідності від дослідних

Результати обчислень за формулами (4.8) для довірчої імовірності  $P = 0,9$  (при  $t = 1,64$ ) наведені в комірках G 44...I 45 розрахункової таблиці з рисунка 6.14. Середні значення й довірчі межі відносних відхилень показують, що формула В.П. Некрасова (6.14) найкраще відображає теплопровідність легких бетонів у нормальних умовах експлуатації, для яких вона дає достатні, але не надмірні запаси (у середньому 10,4%).

Для вологих умов формула (6.14) не забезпечує достатнього запасу надійності. Від'ємне середнє відхилення  $-5,2\%$  показує, що теоретична формула систематичне занижує значення теплопровідності. Для сухого матеріалу формула (6.14) дає надто великі запаси: середнє відхилення дорівнює  $+38,3\%$ , а верхня довірча межа  $+80,3\%$ , тобто реальні значення коефіцієнтів теплопровідності можуть завищуватися майже удвічі.

Широкий довірчий інтервал відхилень, який може сягати 83%, обумовлено не лише недосконалістю формули (6.14), але й значним випадковим розкидом дослідних даних, який видно з рисунків 6.15 і 6.16. При одних і тих же значеннях густини експериментальні значення теплопровідності бетонів різного виду можуть відрізнятися до 40%.

## 6.8. Порівняння статистичних характеристик вибірок

Необхідно порівняти вибірки річних максимумів ваги снігового покриву, отримані за результатами снігомірних зйомок на метеостанціях міст Кропивницький та Знам'янка, з метою об'єднання цих даних в одну вибірку більшого обсягу.

Значення ваги снігового покриву протягом 20 років спостережень, виражені в паскалях, занесені в комірки В 3...С 22 робочого аркуша "6.8" файлу "Аналіз\_даних.xlsx", копія якого зображена на рисунку 6.18. Згідно з рекомендаціями [3], з даних вилучені нульові значення ваги снігового покриву, що відповідають безсніжним рокам. У результаті отримані вибірки обсягом  $N = 18$  даних для метеостанції Кропивницький та  $N = 16$  даних для метеостанції Знам'янка. Результати статистичної обробки обох вибірок за методикою підрозділу 3.3 з використанням статистичних функцій, описаних в підрозділі 3.6, наведені в комірках В 24...С 30 робочого аркуша. Перегляд вихідних даних вказує на певну синхронність змін снігового покриву в часі та на близькість статистичних характеристик обох вибірок для метеостанцій, віддалених приблизно на 30 км.

	A	B	C	D	E	F	G	H	I	J	K
1		<b>Вага снігового покриву</b>									
2		Кропивницький	Знам'янка		<b>Порівняння математичних сподівань</b>						
3		696	748		Z=	0,0298					
4					Alfa=	0,05					
5		77	182		Zcr=	1,96					
6		1683	1189								
7		228									
8			209		<b>Порівняння дисперсій</b>						
9		396	304		F=	1,0097					
10		380	462		k1=	17					
11		340	255		k2=	15					
12		114			Alfa=	0,05					
13		798	528		Fcr=	2,368					
14		841	540								
15		650	480								
16		408	308		Однофакторный дисперсионный анализ						
17		75	143								
18		175			ИТОГИ						
19		255	432		Группы	Счет	Сумма	Среднее	Дисперсия		
20		992	1664		Столбец 1	18	9088	504,8889	161681,6		
21		360	280		Столбец 2	16	8144	509	160135,7		
22		620	420								
23											
24	<b>N=</b>	18	16		Дисперсионный анализ						
25	<b>Min=</b>	75	143		Источник вари	SS	df	MS	F	λ-Значение критическое	
26	<b>Max=</b>	1683	1664		Между гр	143,1634	1	143,1634	0,000889	0,976393	4,149097
27	<b>M=</b>	505	509		Внутри гр	5150624	32	160957			
28	<b>S=</b>	402	400								
29	<b>V=</b>	0,796	0,786		Итого	5150767	33				
30	<b>D=</b>	161682	160136								

Рис. 6.18. Порівняння вибірок ваги снігового покриву на метеостанціях Кропивницький і Знам'янка

**Порівняння математичних сподівань** виконане за методикою підрозділу 5.1. За формулою (5.2) в комірці F 3 робочого аркуша обчислене значення статистики критерію

$$Z = |505 - 509| / \sqrt{\frac{161682}{18} + \frac{160136}{16}} = 0,0298 < Z_{cr} = 1,96$$

Критичне значення критерію  $Z_{cr}=1,96$  визначене за таблицею з підрозділу 5.1 для рівня значимості 0,05. Оскільки  $Z < Z_{cr}$ , математичні сподівання річних максимумів ваги снігового покриву на метеостанціях міст Кропивницький та Знамянка можна вважати практично рівними.

**Порівняння дисперсій** тих же вибірок здійснене за описаним у підрозділі 5.1 критерієм Фішера. Більшою отримана дисперсія вибірки річних максимумів ваги снігового покриву на метеостанції Кропивницький, тому цю вибірку позначаємо номером 1. З результатів статистичної обробки вибираємо необхідні характеристики:

$$\begin{aligned} M_1 &= 505 \text{ Па}, & D_1 &= 161682 \text{ Па}^2, & N_1 &= 18 \text{ даних}, \\ M_2 &= 509 \text{ Па}, & D_2 &= 160136 \text{ Па}^2, & N_2 &= 16 \text{ даних}. \end{aligned}$$

Статистика критерію Фішера обчислена в комірці F 9 за формулою (5.3):  $F = D_1/D_2 = 161682/160136 = 1,0097$ .

В комірках F 10 та F 11 розрахункового бланка обчислені кількості ступенів свободи  $k_1 = N_1 - 1 = 17$ ,  $k_2 = N_2 - 1 = 15$ , а в комірці F 10 з використанням функції  $F.OBP.PX(F12;F10;F11)$  визначене критичне значення критерію  $F_{cr} = 2,368$  за цими ступенями свободи та заданим у комірці F 12 рівнем значимості  $\alpha = 0,05$ .

У результаті виконаних обчислень отримано  $F = 1,0097 < F_{cr} = 2,368$ , що дозволяє вважати дисперсії порівняних вибірок практично рівними.

**Дисперсійний аналіз** для двох описаних вище вибірок виконано засобами Microsoft Excel згідно з рекомендаціями підрозділу 5.1. Для запуску процедури дисперсійного аналізу слід виконати такі операції:

- увійти у вкладку "Данные" та обрати групу команд "Анализ";
- увійти в блок "Анализ данных" та обрати команду "Однофакторный дисперсионный анализ";
- обвести курсором миші вхідний інтервал в комірках В 3...С 22;
- встановити групування по стовпцях;
- обрати вихідний інтервал (відмітити мишею ліву верхню комірку області, у яку буде виведена таблиця результатів – Е 16);
- задати рівень значимості 0,05 і натиснути ОК.

Результати дисперсійного аналізу, автоматично виведені в комірки E 16...K 29, містять таку інформацію:

- обсяги кожної з вибірок, суми даних, середні значення та дисперсії в комірках E 19...I 21;
- міжгрупові (факторні) та внутрішньогрупові (залишкові) показники;
- дослідне значення критерію Фішера  $F = 0,976$  в комірці J 26 та критичне значення  $F_{cr} = 4,149$  в комірці K 26;

Оскільки в результаті обчислень отримано  $F = 0,976 < F_{cr} = 4,149$ , порівняні вибірки можна вважати однорідними.

Таким чином, результати трьох виконаних перевірок (порівняння математичних сподівань, порівняння дисперсій, дисперсійний аналіз) вказують на однорідність вибірок річних максимумів ваги снігового покриву на метеостанціях Кропивницький і Знам'янка. Ці вибірки можна об'єднати з метою збільшення обсягу даних та імовірнісного опису снігового навантаження на території, що охоплює ці міста.

### 6.9. Нелінійне перетворення випадкової величини методом лінеаризації

Нижче наведено приклад використання описаного в підрозділі 5.3 на основі підручника [2] методу лінеаризації для визначення числових характеристик результату нелінійного перетворення випадкової величини швидкості вітру  $v$  у вітровий тиск  $w$ . Числові значення статистичних характеристик випадкової величини швидкості вітру приймемо за даними прикладу з підрозділу 6.3 рівними  $M_v = 3,97$  м/с та  $S_v = 2,21$  м/с.

Відповідно до відомої формули (5.14), загальна формула (5.17) нелінійного квадратичного перетворення однієї випадкової величини набуває вигляду

$$w = \varphi(v) = 0,61 v^2 . \quad (6.15)$$

Згідно з (5.18), математичне сподівання вітрового тиску дорівнює

$$M_w = 0,61 M_v^2 = 0,61 \times 3,97^2 = 9,6 \text{ Па} . \quad (6.16)$$

Похідна функції (6.15) по аргументу  $v$  дорівнює

$$\frac{\partial v}{\partial w} = 1,22 v . \quad (6.17)$$

Підставивши (6.17) до загальної формули (5.19) та замінивши змінну величину  $v$  на її математичне сподівання  $M_v$ , як це вказано в підрозділі 5.3, отримаємо вираз для дисперсії вітрового тиску. Корінь квадратний з цього виразу дає робочу формулу для обчислення стандарту вітрового тиску у такому вигляді:

$$S_w = \frac{\partial v}{\partial w} S_v = 1,22 M_v S_v = 1,22 \times 3,97 \times 2,21 = 10,7 \text{ Па} . \quad (6.18)$$

Для обчислення статистичних характеристик вітрового тиску при інших значеннях математичного сподівання й стандарту швидкості вітру слід скористатися формулами (6.16) та (6.18).

Порівняємо коефіцієнти варіації розподілу швидкості вітру та вітрового тиску, які дорівнюють

$$V_v = \frac{S_v}{M_v} = \frac{2,21}{3,97} = 0,557, \quad V_w = \frac{S_w}{M_w} = \frac{10,7}{9,6} = 1,115. \quad (6.19)$$

Порівняння отриманих величин (6.19) показує, що в результаті квадратичного перетворення швидкості в тиск вітру коефіцієнт варіації збільшився у 2,002 рази. Це співвідношення цілком узгоджується з даними монографії [3], де за результатами статистичних досліджень вітрового навантаження на метеостанціях України показано, що розподіли поточних значень швидкості вітру описуються законом Вейбулла з пагорбоподібною формою та коефіцієнтом варіації  $V_v < 1$ . Розподіли вітрового тиску також описується законом Вейбулла, але з експонентоподібною формою та удвічі більшим коефіцієнтом варіації  $V_w = 2 \times V_v > 1$ . Відповідність відношення коефіцієнтів варіації швидкості та тиску вітру до раніше отриманих даних [3] підтверджує адекватність перетворень, виконаних у даному прикладі.

## РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Барковський В.В., Барковська Н.В., Лопатін О.К. Теорія ймовірностей та математична статистика. 5-те видання. – Київ: Центр учбової літератури, 2010. – 424 с.
2. Глеч С.Г. Теорія ймовірностей та математична статистика: навч. посібник/ С.Г. Глеч, С.Ф. Ледаєв, І.В. Ольшанська. – Севастополь: СевНТУ, 2011. – 176 с. ISBN 978 – 617 – 612 – 005 – 6
3. Пашинський В.А. Атмосферні навантаження на будівельні конструкції на території України. К.: УкрНДІпроектстальконструкція, 1999.- 185 с.
4. Теорія ймовірностей та математична статистика: Навчальний посібник / О.І. Кушлик-Дивульська, Н. В. Поліщук, Б. П. Орел, П.І. Штабалюк. – К: НТУУ «КПІ», 2014. – 212 с.
5. Центр довідки з Excel. [Електронний ресурс] Режим доступу: <https://support.office.com/uk-UA/Excel>
6. Приклади виконання статистичного аналізу даних у середовищі Microsoft Excel у файлі **Аналіз\_даних.xlsx**.

## Додаток А

### СТАТИСТИЧНІ ТАБЛИЦІ

Таблиця А.1. Функція нормального розподілу  
(наведені лише десяткові знаки після коми)

t	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	50000	50398	50797	51196	51595	51993	52392	52790	53188	53585
0,1	53982	54379	54775	55171	55567	55961	56355	56749	57142	57534
0,2	57925	58316	58706	59095	59483	59870	60256	60641	61026	61409
0,3	61791	62171	62551	62930	63307	63683	64057	64430	64802	65173
0,4	65542	65909	66275	66640	67003	67364	67724	68082	68438	68793
0,5	69146	69497	69846	70194	70540	70884	71226	71566	71904	72240
0,6	72574	72906	73237	73565	73891	74215	74537	74857	75174	75490
0,7	75803	76114	76423	76730	77035	77337	77637	77935	78230	78523
0,8	78814	79103	79389	79673	79954	80233	80510	80784	81057	81326
0,9	81594	81858	82121	82381	82639	82894	83147	83397	83645	83891
1,0	84134	84375	84613	84849	85083	85314	85542	85769	85992	86214
1,1	86433	86650	86864	87076	87285	87492	87697	87899	88099	88297
1,2	88493	88686	88876	89065	89251	89435	89616	89795	89972	90147
1,3	90319	90490	90658	90824	90987	91149	91308	91465	91620	91773
1,4	91924	92073	92219	92364	92506	92647	92785	92921	93056	93188
1,5	93319	93447	93574	93699	93821	93942	94062	94179	94294	94408
1,6	94520	94630	94738	94844	94949	95052	95154	95254	95352	95448
1,7	95543	95636	95728	95818	95907	95994	96079	96163	96246	96327
1,8	96406	96485	96562	96637	96711	96784	96855	96925	96994	97062
1,9	97128	97193	97257	97319	97381	97441	97500	97558	97614	97670
2,0	97724	97778	97830	97882	97932	97981	98030	98077	98123	98169
2,1	98213	98257	98299	98341	98382	98422	98461	98499	98537	98573
2,2	98609	98644	98679	98712	98745	98777	98808	98839	98869	98898
2,3	98927	98955	98982	99009	99035	99061	99086	99110	99134	99157
2,4	99180	99202	99223	99245	99265	99285	99305	99324	99343	99361
2,5	99379	99396	99413	99429	99445	99461	99476	99491	99505	99520
2,6	99533	99547	99560	99573	99585	99597	99609	99620	99631	99642
2,7	99653	99663	99673	99683	99692	99702	99710	99719	99728	99736
2,8	99744	99752	99759	99767	99774	99781	99788	99794	99801	99807
2,9	99813	99819	99824	99830	99835	99841	99846	99851	99855	99860
3,0	99865	99869	99873	99877	99881	99885	99889	99892	99896	99899
3,1	99903	99906	99909	99912	99915	99918	99921	99923	99926	99928
3,2	99931	99933	99935	99938	99940	99942	99944	99946	99948	99949
3,3	99951	99953	99954	99956	99958	99959	99961	99962	99963	99965
3,4	99966	99967	99968	99969	99970	99971	99972	99973	99974	99975
3,5	99976	99977	99978	99979	99979	99980	99981	99982	99982	99983
3,6	99984	99984	99985	99985	99986	99986	99987	99987	99988	99988
3,7	99989	99989	99990	99990	99990	99991	99991	99991	99992	99992
3,8	99992	99993	99993	99993	99993	99994	99994	99994	99994	99994
3,9	99995	99995	99995	99995	99995	99996	99996	99996	99996	99996

Таблиця А.2. Критичні значення критерію узгодженості Пірсона

Ступені свободи $k$	Рівні значимості $\alpha$				
	0,20	0,10	0,05	0,02	0,01
1	1,64	2,71	3,84	5,41	6,64
2	3,22	4,60	5,99	7,82	9,21
3	4,64	6,25	7,82	9,84	11,3
4	5,99	7,78	9,49	11,7	13,28
5	7,29	9,24	11,1	13,4	15,1
6	8,56	10,6	12,6	15,0	16,8
7	9,80	12,0	14,1	16,6	18,5
8	11,0	13,4	15,5	18,2	20,1
9	12,2	14,7	16,9	19,7	21,7
10	13,4	16,0	18,3	21,2	23,2
11	14,6	17,3	19,7	22,6	24,7
12	15,8	18,5	21,0	24,1	26,2
13	16,9	19,8	22,4	25,5	27,7
14	18,2	21,1	23,7	26,9	29,1
15	19,3	22,3	25,0	28,3	30,6
16	20,5	23,5	26,3	29,6	32,0
17	21,6	24,8	27,6	31,0	33,4
18	22,8	26,0	28,9	32,3	34,8
19	23,9	27,2	30,1	33,7	36,2
20	25,0	28,4	31,4	35,0	37,6
21	26,2	29,6	32,7	36,3	38,9
22	27,3	30,8	33,9	37,7	40,3
23	28,4	32,0	35,2	39,0	41,6
24	29,6	33,2	36,4	40,3	43,0
25	30,3	34,4	37,6	41,7	44,3
26	31,8	35,6	38,9	42,9	45,6
27	32,9	36,7	40,1	44,1	47,0
28	34,0	37,9	41,3	45,4	48,3
29	35,1	39,1	42,6	46,7	49,6
30	36,2	40,3	43,8	48,0	50,9
35	41,8	46,1	49,8	53,2	57,3
40	47,3	51,8	55,8	59,3	63,7

Таблиця А.3. Критичні значення критерію Фішера для  $\alpha=0,01$ 

k <sub>2</sub>	Кількість ступенів свободи k <sub>1</sub>														
	1	2	3	4	5	6	8	10	12	15	20	25	30	40	50
1	4052	5000	5403	5625	5764	5859	5981	6056	6106	6157	6209	6240	6261	6287	6303
2	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,5	27,2	27,1	26,9	26,7	26,6	26,5	26,4	26,4
4	21,2	18,0	16,7	16,0	15,5	15,2	14,8	14,5	14,4	14,2	14,0	13,9	13,8	13,7	13,7
5	16,3	13,3	12,1	11,4	11,0	10,7	10,3	10,1	9,89	9,72	9,55	9,45	9,38	9,29	9,24
6	13,7	10,9	9,78	9,15	8,75	8,47	8,10	7,87	7,72	7,56	7,40	7,30	7,23	7,14	7,09
8	11,3	8,65	7,59	7,01	6,63	6,37	6,03	5,81	5,67	5,52	5,36	5,26	5,20	5,12	5,07
10	10,0	7,56	6,55	5,99	5,64	5,39	5,06	4,85	4,71	4,56	4,41	4,31	4,25	4,17	4,12
12	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,30	4,16	4,01	3,86	3,76	3,70	3,62	3,57
15	8,68	6,36	5,42	4,89	4,56	4,32	4,00	3,80	3,67	3,52	3,37	3,28	3,21	3,13	3,08
20	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,37	3,23	3,09	2,94	2,84	2,78	2,69	2,64
25	7,77	5,57	4,68	4,18	3,85	3,63	3,32	3,13	2,99	2,85	2,70	2,60	2,54	2,45	2,40
30	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,98	2,84	2,70	2,55	2,45	2,39	2,30	2,25
40	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,80	2,66	2,52	2,37	2,27	2,20	2,11	2,06
50	7,17	5,06	4,20	3,72	3,41	3,19	2,89	2,70	2,56	2,42	2,27	2,17	2,10	2,01	1,95

Таблиця А.4. Критичні значення критерію Фішера для  $\alpha=0,02$ 

k <sub>2</sub>	Кількість ступенів свободи k <sub>1</sub>														
	1	2	3	4	5	6	8	10	12	15	20	25	30	40	50
1	1013	1250	1351	1406	1441	1464	1495	1514	1526	1539	1552	1560	1565	1571	1575
2	48,5	49,0	49,2	49,2	49,3	49,3	49,4	49,4	49,4	49,4	49,4	49,5	49,5	49,5	49,5
3	20,6	18,9	18,1	17,7	17,4	17,2	17,0	16,9	16,8	16,7	16,6	16,5	16,4	16,4	16,4
4	14,0	12,1	11,3	10,9	10,6	10,4	10,2	10,0	9,89	9,78	9,67	9,60	9,55	9,50	9,46
5	11,3	9,45	8,67	8,23	7,95	7,76	7,50	7,34	7,23	7,12	7,01	6,94	6,89	6,83	6,80
6	9,88	8,05	7,29	6,86	6,58	6,39	6,14	5,98	5,88	5,76	5,65	5,58	5,53	5,47	5,44
8	8,39	6,64	5,90	5,49	5,22	5,04	4,79	4,63	4,53	4,42	4,30	4,23	4,19	4,13	4,09
10	7,64	5,93	5,22	4,82	4,55	4,37	4,13	3,97	3,87	3,76	3,64	3,57	3,52	3,46	3,43
12	7,19	5,52	4,81	4,42	4,16	3,98	3,74	3,59	3,48	3,37	3,25	3,18	3,13	3,07	3,03
15	6,77	5,14	4,45	4,06	3,81	3,63	3,39	3,23	3,13	3,02	2,90	2,83	2,78	2,71	2,67
20	6,39	4,79	4,11	3,73	3,48	3,30	3,07	2,91	2,81	2,70	2,58	2,50	2,45	2,38	2,34
25	6,18	4,59	3,93	3,55	3,30	3,13	2,89	2,74	2,63	2,52	2,40	2,32	2,27	2,20	2,16
30	6,04	4,47	3,81	3,43	3,19	3,01	2,78	2,62	2,52	2,40	2,28	2,20	2,15	2,08	2,04
40	5,87	4,32	3,67	3,30	3,05	2,88	2,64	2,49	2,38	2,26	2,14	2,06	2,01	1,93	1,89
50	5,78	4,23	3,59	3,22	2,97	2,80	2,56	2,41	2,30	2,18	2,06	1,98	1,92	1,85	1,80

Таблиця А.5. Критичні значення критерію Фішера для  $\alpha=0,05$ 

$k_2$	Кількість ступенів свободи $k_1$														
	1	2	3	4	5	6	8	10	12	15	20	25	30	40	50
1	161	200	216	225	230	234	239	242	244	246	248	249	250	251	252
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,85	8,79	8,74	8,70	8,66	8,63	8,62	8,59	8,58
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,70
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,68	4,62	4,56	4,52	4,50	4,46	4,44
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	4,00	3,94	3,87	3,83	3,81	3,77	3,75
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,35	3,28	3,22	3,15	3,11	3,08	3,04	3,02
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,98	2,91	2,85	2,77	2,73	2,70	2,66	2,64
12	4,75	3,89	3,49	3,26	3,11	3,00	2,85	2,75	2,69	2,62	2,54	2,50	2,47	2,43	2,40
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,54	2,48	2,40	2,33	2,28	2,25	2,20	2,18
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35	2,28	2,20	2,12	2,07	2,04	1,99	1,97
25	4,24	3,39	2,99	2,76	2,60	2,49	2,34	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,84
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,16	2,09	2,01	1,93	1,88	1,84	1,79	1,76
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,08	2,00	1,92	1,84	1,78	1,74	1,69	1,66
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	2,03	1,95	1,87	1,78	1,73	1,69	1,63	1,60

Таблиця А.6. Критичні значення критерію Фішера для  $\alpha=0,10$ 

$k_2$	Кількість ступенів свободи $k_1$														
	1	2	3	4	5	6	8	10	12	15	20	25	30	40	50
1	39,9	49,5	53,6	55,8	57,2	58,2	59,4	60,2	60,7	61,2	61,7	62,1	62,3	62,5	62,7
2	8,53	9,00	9,16	9,24	9,29	9,33	9,37	9,39	9,41	9,42	9,44	9,45	9,46	9,47	9,47
3	5,54	5,46	5,39	5,34	5,31	5,28	5,25	5,23	5,22	5,20	5,18	5,17	5,17	5,16	5,15
4	4,54	4,32	4,19	4,11	4,05	4,01	3,95	3,92	3,90	3,87	3,84	3,83	3,82	3,80	3,80
5	4,06	3,78	3,62	3,52	3,45	3,40	3,34	3,30	3,27	3,24	3,21	3,19	3,17	3,16	3,15
6	3,78	3,46	3,29	3,18	3,11	3,05	2,98	2,94	2,90	2,87	2,84	2,81	2,80	2,78	2,77
8	3,46	3,11	2,92	2,81	2,73	2,67	2,59	2,54	2,50	2,46	2,42	2,40	2,38	2,36	2,35
10	3,29	2,92	2,73	2,61	2,52	2,46	2,38	2,32	2,28	2,24	2,20	2,17	2,16	2,13	2,12
12	3,18	2,81	2,61	2,48	2,39	2,33	2,24	2,19	2,15	2,10	2,06	2,03	2,01	1,99	1,97
15	3,07	2,70	2,49	2,36	2,27	2,21	2,12	2,06	2,02	1,97	1,92	1,89	1,87	1,85	1,83
20	2,97	2,59	2,38	2,25	2,16	2,09	2,00	1,94	1,89	1,84	1,79	1,76	1,74	1,71	1,69
25	2,92	2,53	2,32	2,18	2,09	2,02	1,93	1,87	1,82	1,77	1,72	1,68	1,66	1,63	1,61
30	2,88	2,49	2,28	2,14	2,05	1,98	1,88	1,82	1,77	1,72	1,67	1,63	1,61	1,57	1,55
40	2,84	2,44	2,23	2,09	2,00	1,93	1,83	1,76	1,71	1,66	1,61	1,57	1,54	1,51	1,48
50	2,81	2,41	2,20	2,06	1,97	1,90	1,80	1,73	1,68	1,63	1,57	1,53	1,50	1,46	1,44

## З М І С Т

ВСТУП .....	3
1. БАЗОВІ ПОНЯТТЯ ТЕОРІЇ ІМОВІРНОСТЕЙ І МАТЕМАТИЧНОЇ СТАТИСТИКИ .....	4
2. ВИПАДКОВІ ВЕЛИЧИНИ ТА ЇХ ХАРАКТЕРИСТИКИ .....	7
2.1. Основні властивості випадкових величин .....	7
2.2. Закон розподілу неперервної випадкової величини .....	8
2.3. Числові характеристики випадкової величини .....	10
2.4. Поширені закони розподілу випадкових величин .....	14
2.4.1. Закон рівномірного розподілу .....	14
2.4.2. Нормальний закон розподілу .....	15
2.4.3. Логнормальний закон розподілу .....	16
2.4.4. Закон розподілу Гумбеля .....	17
2.4.5. Закон розподілу Вейбулла .....	18
2.4.6. Експоненціальний закон розподілу .....	20
2.4.7. Зрізані розподіли .....	20
2.5. Операції з випадковими величинами .....	22
3. СТАТИСТИЧНА ОБРОБКА ВИБІРОК ВИПАДКОВИХ ВЕЛИЧИН .....	24
3.1. Загальний порядок формування і статистичної обробки вибірки випадкової величини .....	24
3.2. Побудова гістограми розподілу вибірки .....	27
3.3. Визначення числових характеристик вибірки .....	30
3.4. Точність оцінювання числових характеристик .....	33
3.5. Вибір закону розподілу за результатами статистичного аналізу дослідних даних .....	37
3.6. Математичне забезпечення статистичної обробки вибірок випадкових величин в середовищі Microsoft Excel .....	42
4. АПРОКСИМАЦІЯ ЕМПІРИЧНИХ ЗАЛЕЖНОСТЕЙ .....	44
4.1. Кореляція та лінійна регресія випадкових величин .....	44
4.2. Опис нелінійних залежностей типовими функціями .....	47
4.3. Опис нелінійних та багатofакторних залежностей довільними функціями .....	50
4.4. Узгодженість теоретичних результатів з дослідними даними .....	52
4.5. Математичне забезпечення апроксимації емпіричних залежностей в середовищі Microsoft Excel .....	54
5. БАГАТОВИМІРНІ СТАТИСТИЧНІ МЕТОДИ .....	56

5.1. Порівняння випадкових величин та дисперсійний аналіз.....	56
5.2. Багатофакторні експерименти та їх аналіз.....	59
5.3. Функції випадкових величин.....	62
5.4. Поняття про випадкові процеси.....	65
5.5. Математичне забезпечення багатовимірною статистичного аналізу в середовищі Microsoft Excel.....	67
6. ПРИКЛАДИ ЗАСТОСУВАННЯ СТАТИСТИЧНИХ МЕТОДІВ ПРИ ВИКОНАННІ ІНЖЕНЕРНИХ ДОСЛІДЖЕНЬ.....	69
6.1. Аналіз середньої густини пінополістиролу.....	69
6.2. Дослідження та нормування снігового навантаження.....	71
6.3. Аналіз даних по швидкості вітру.....	74
6.4. Апроксимація лінійних залежностей.....	77
6.5. Апроксимація нелінійної залежності типовими функціями.....	81
6.6. Апроксимація двофакторної залежності.....	85
6.7. Порівняння теоретичних результатів з дослідними даними.....	90
6.8. Порівняння статистичних характеристик вибірок.....	95
6.9. Нелінійне перетворення випадкової величини методом лінеаризації.....	97
РЕКОМЕНДОВАНА ЛІТЕРАТУРА.....	99
Додаток А Статистичні таблиці.....	100
Таблиця А.1 Функція нормального розподілу.....	100
Таблиця А.2 Критичні значення критерію узгодженості Пірсона.....	101
Таблиця А.3 Критичні значення критерію Фішера для $\alpha=0,01$ .....	102
Таблиця А.4 Критичні значення критерію Фішера для $\alpha=0,02$ .....	102
Таблиця А.5 Критичні значення критерію Фішера для $\alpha=0,05$ .....	103
Таблиця А.6 Критичні значення критерію Фішера для $\alpha=0,10$ .....	103
Додаток Б Приклади виконання статистичного аналізу даних у середовищі Microsoft Excel, наведені в окремому файлі <b>Аналіз_даних.xlsx</b>	

Навчальне видання

Пашинський В.А., Пашинський М.В.

**СТАТИСТИЧНІ МЕТОДИ  
В ІНЖЕНЕРНИХ ДОСЛІДЖЕННЯХ**

Навчальний посібник для здобувачів вищої освіти  
з інженерних спеціальностей

Електронне видання

Укладачі – доктор технічних наук, професор  
Пашинський Віктор Антонович,  
кандидат технічних наук  
Пашинський Микола Вікторович