Міністерство освіти і науки України
Національний університет водного господарства та природокористування
Навчально-науковий інститут агроекології та землеустрою
Кафедра геодезії та картографії

05-04-124М

МЕТОДИЧНІ ВКАЗІВКИ

до виконання практичних та самостійних робіт
з навчальної дисципліни «Математична обробка геодезичних вимірів»
для здобувачів вищої освіти першого (бакалаврського) рівня
за освітньо-професійною програмою «Геодезія та землеустрій»
спеціальності 193 «Геодезія та землеустрій»
денної та заочної форм навчання

ЕЛЕМЕНТИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

Рекомендовано науково-методичною
радою з якості ННІАЗ
Протокол № 10 від 25.04.2023 р.

Рівне – 2023

Методичні вказівки до виконання практичних та самостійних робіт з навчальної дисципліни «Математична обробка геодезичних вимірів» для здобувачів вищої освіти першого (бакалаврського) рівня за освітньо-професійною програмою «Геодезія та землеустрій» спеціальності 193 «Геодезія та землеустрій» денної та заочної форм навчання. Елементи математичної статистики [Електронне видання] / Тадєєв О. А. – Рівне : НУВГП, 2023. – 46 с.

Укладач: Тадєєв О. А., доцент кафедри геодезії та картографії, кандидат технічних наук, доцент.

Відповідальний за випуск: Янчук Р. М., завідувач кафедри геодезії та картографії, кандидат технічних наук, доцент.

Керівник групи забезпечення спеціальності 193 «Геодезія та землеустрій»: доктор сільськогосподарських наук, професор Мошинський В. С.

© О. А. Тадєєв, 2023
© НУВГП, 2023

Зміст

Основні терміни та визначення математичної статистики.....3
1. Визначення закону розподілу випадкової величини на основі дослідних даних.....9
2. Визначення точності та надійності числових оцінок параметрів розподілу в умовах обмеженої кількості випробувань…….....22
3. Визначення тісноти та форми кореляційного зв'язку в системі двох випадкових величин.…….….....29
Література…….....38
Додатки…….....39

ОСНОВНІ ТЕРМІНИ ТА ВИЗНАЧЕННЯ МАТЕМАТИЧНОЇ СТАТИСТИКИ

В основі всіх законів розподілу випадкових величин та їх числових характеристик лежить дослідження. Кожне дослідження випадкових явищ, яке ґрунтується на теорії ймовірностей, прямо або опосередковано опирається на експериментальні дані. Теорія ймовірностей дає змогу теоретичним шляхом визначати ймовірності одних подій за ймовірностями інших, закони розподілу і числові характеристики одних випадкових величин за законами розподілу й числовими характеристиками інших. Такий опосередкований підхід дає змогу значно заощадити час і засоби, які витрачаються на експеримент, але не виключають самого експерименту. Кожне дослідження в області випадкових явищ так чи інакше завжди опирається на систему спостережень, експеримент, експериментальні дані.

Розробка методів реєстрації, накопичування, описування та аналізу статистичних експериментальних даних для дослідження масових випадкових явищ складають **предмет математичної статистики**.

Оперуючи наборами експериментальних даних, математична статистика розв'язує наступні основні завдання (див. схему на рис. 1):

– вирівнювання статистичного ряду розподілу;
– перевірка правдоподібності статистичних гіпотез;
– визначення невідомих параметрів розподілу випадкової величини.

Окреслені завдання математичної статистики вирішуються як осібно, так і в комплексі. Наприклад, вирішення завдання визначення закону розподілу випадкової величини – це комплексне завдання, яке передбачає вирівнювання статистичного ряду розподілу випадкової величини і за його результатами перевірку гіпотези про підпорядкованість статистичного розподілу тому чи іншому теоретичному законові.

Всі завдання зводяться до обробки результатів спостережень масових випадкових явищ. Залежно від характеру та змісту поставленої задачі та об'єму експериментальних даних завдання набувають різних форм. Різні форми завдань формують математичний апарат, завдяки якому у конкретних практичних умовах реалізовуються теореми й закони теорії ймовірностей.

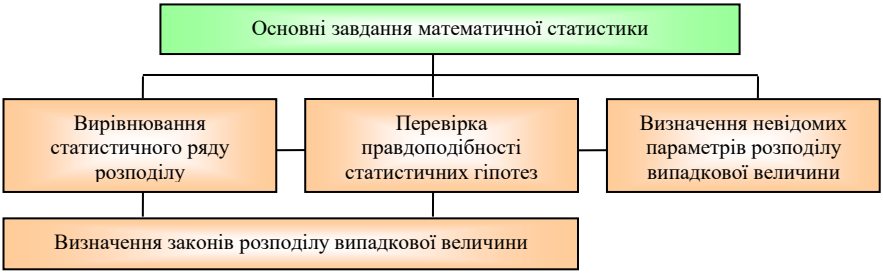


Рис. 1. Основні завдання математичної статистики

Допустимо, досліджується деяка випадкова величини *X*. З цією метою над величиною проводиться ряд незалежних випробувань (спостережень, вимірів, дослідів), у кожному з яких вона набуває того чи іншого значення. Сукупність одержаних значень є первинним матеріалом, який підлягає подальшій статистичній обробці та аналізу. Така сукупність результатів випробувань називається **проста статистична (або генеральна) сукупність** – це таблиця, у першому стовпці якої зазначено номери випробувань *i*, у наступному – результати випробувань x_i (табл. 1).

Таблиця 1

Проста статистична сукупність

Номери випробувань *i*	Результати випробувань x_i
1	x_1
2	x_2
…	…
n	x_n

За умови достатньо великої кількості результатів випробувань (n>100) проста статистична сукупність буде надмірно громіздкою і перестає бути зручною формою представлення первинного статистичного матеріалу. Тоді просту статистичну сукупність перетворюють у більш компактну форму. У ході перетворень дотримуються наступної черговості дій:

1. Весь діапазон набутих у результаті випробувань значень випадкової величини поділяють на k інтервалів I_i ($i=\overrightarrow{1,k}$).
2. Вираховують числа m_i ($i=\overrightarrow{1,k}$) значень величини, які потрапляють у межі кожного інтервалу.
3. Обчислюють відносні частоти (статистичні ймовірності) p_i^* попадання значень величини у межі кожного інтервалу, для чого відповідні інтервалам числа m_i ділять на загальну кількість випробувань n :

$$p_i^* = \frac{m_i}{n}.$$

Сума обчислених відносних частот повинна дорівнювати одиниці: $\sum_{i=1}^{k} p_i^* = 1$.

Одержані у ході перетворень результати поміщають до таблиці, яка називається **статистичний ряд розподілу** – це таблиця, у першому рядку якої зазначено інтервали у порядку їх розташування вздовж осі абсцис, у другому рядку – відповідні інтервалам відносні частоти (табл. 2).

Таблиця 2

Статистичний ряд розподілу

I_i	$x_1'; x_2'$	$x_2'; x_3'$	…	$x_k'; x_{k+1}'$
p_i^*	p_1^*	p_2^*	…	p_k^*

В табл. 2 позначено x_i' – границі інтервалів. У ході побудови статистичного ряду розподілу дотримуються наступних рекомендацій:

1. Кількість інтервалів k доцільно вибирати порядку 10 і більше при відповідно зростаючому об'ємі статистичного матеріалу. Проте якщо k надмірно велике, то ряд стає невиразним і частоти набувають незакономірних коливань. Разом з тим, потрібно забезпечувати вимогу, щоб значення m_i були достатньо репрезентативними з імовірнісної точки зору. Для цього рекомендується умова, щоб $m_i > 5$.
2. Довжини інтервалів не обов'язково повинні бути рівними поміж собою. У тих частинах загального діапазону значень величини, де вони зосереджені з більшою щільністю і значення m_i відповідно більші за величиною, рекомендується обирати інтервали більш вузькими порівняно з іншими.
3. Якщо величина набула значення, яке дорівнює одній з границь інтервалів x_i', то рекомендується вважати це значення належним рівною мірою до обох суміжних інтервалів і додавати до чисел m_i одного й іншого інтервалів по 0,5. Допускається також, з метою однозначності, такі значення відносити до тих інтервалів, у яких ліва границя дорівнює відповідному значенню величини.

Гістограма – це графічне оформлення статистичного ряду розподілу (див. рис. 2). Для побудови гістограми на осі абсцис позначають інтервали статистичного ряду розподілу. На кожному інтервалі, беручи його за основу, будують прямокутник з площею, яка дорівнює відносній частоті інтервалу p_i^*. Для цього, завчасно задавшись масштабом, від основи кожного прямокутника відкладають його висоту, яка є результатом ділення відносної частоти інтервалу на його довжину. Якщо інтервали рівні за довжиною, то

висоти прямокутників пропорційні відповідним частотам. Беручи до уваги, що $\sum_{i=1}^{k} p_i^* = 1$, загальна площа гістограми дорівнює одиниці.

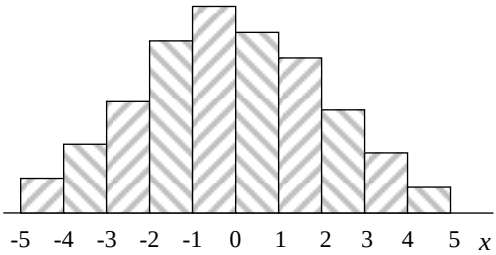


Рис. 2. Гістограма статистичного розподілу

Проста статистична сукупність чи статистичний ряд розподілу випадкової величини – це первинні форми представлення експериментальних даних для їх подальшої статистичної обробки та аналізу. Одним із способів обробки таких даних є побудова статистичної функції розподілу випадкової величини. **Статистична функція розподілу** $F^*(x)$ випадкової величини X – це відносна частота події $X < x$ у даному статистичному матеріалі:

$$F^*(x) = p^*(X < x).$$

Для визначення функції $F^*(x)$ при заданому значенні змінної x можна використати просту статистичну сукупність. Відтак, достатньо порахувати кількість випробувань $m_{x_i < x}$, в яких величина набула значень, які менші від x, і поділити її на загальну кількість випробувань n:

$$F^*(x) = \frac{m_{x_i < x}}{n}.$$

Статистична функція розподілу – це перервна ступінчата функція з розривами, які відповідають значенням величини і дорівнюють частотам цих значень (рис. 3). $F^*(x)$ достатньо точно представляє дослідні дані. Однак її побудова при великому обсязі статистичного матеріалу перетворюється в надто трудомісткий процес. У таких випадках $F^*(x)$ можна побудувати наближено за даними статистичного ряду розподілу, обираючи значеннями змінної x границі інтервалів:

$$F^*(x) = \sum_{x_i < x} p_i^*.$$

Позначення $x_i < x$ вказує, що для визначення функції при заданій змінній x необхідно додавати частоти всіх інтервалів, у яких ліва границя менша від

значення x. Для побудови графіка наближеної функції $F^*(x)$ вздовж осі абсцис позначають границі інтервалів, вздовж осі ординат – відповідні їм значення функції, а отримані точки з'єднують відрізками прямих (рис. 4).

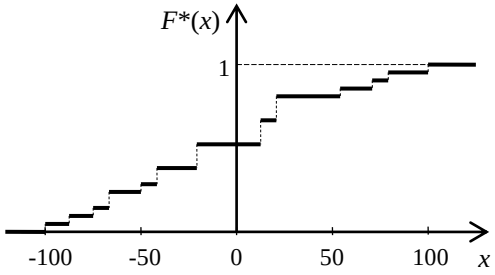


Рис. 3. Графік статистичної функції розподілу

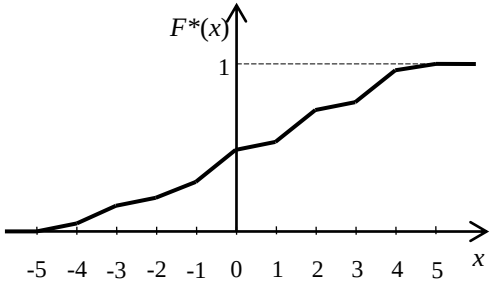


Рис. 4. Графік наближеної статистичної функції розподілу

За умови великої кількості випробувань, гранично $n \to \infty$, згідно теореми Бернуллі закону великих чисел, відносна частота p_i^* події $X < x$ при заданому x необмежено наближається до її ймовірності p. Отже, на цій основі, статистична функція розподілу $F^*(x)$ відповідно гранично буде наближатись до її теоретичного аналога – функції розподілу $F(x)$.

Числові характеристики статистичного розподілу. При описуванні окремих ознак розподілу випадкової величини обчислюють відповідні їм числові характеристики. Математичні характеристики, якими оперують в теорії ймовірностей, використовують також і для статистичних розподілів. Кожна ймовірнісна числова характеристика випадкової величини X має статистичний аналог і зберігає свій зміст.

Для математичного сподівання статистичним аналогом є середнє арифметичне спостережених значень випадкової величини:

$$m_x^* = M^*[X] = \frac{\sum_{i=1}^{n} x_i}{n},$$

де x_i – значення величини, отримане в i-му випробуванні; n – кількість випробувань. Цю характеристику ще називають середнім статистичним. Відповідно до закону великих чисел при $n \to \infty$ середнє арифметичне m_x^* наближається (збігається за ймовірністю) до математичного сподівання m_x. При достатньо великому n середнє арифметичне можна вважати наближено рівним математичному сподіванню. При обмеженій кількості випробувань середнє арифметичне саме є випадковою величиною. Проте m_x^* пов'язане за змістом з m_x і наближено характеризує положення центра статистичного розподілу випадкової величини.

Дисперсія, як характеристика розсіювання значень випадкової величини, – це центральний момент другого порядку (або математичне сподівання квадрата центрованої величини):

$$D_x = \mu_2 = M[\overset{\circ}{X}{}^2] = \sum_{i=1}^{n} (x_i - m_x)^2 p_i .$$

Беручи до уваги можливість заміни математичного сподівання m_x його статистичним аналогом m_x^*, для статистичної дисперсії одержимо:

$$D_x^* = \frac{\sum_{i=1}^{n} (x_i - m_x^*)^2}{n}.$$

Аналогічні заміни визначають статистичні початкові та центральні моменти будь-якого порядку s:

$$\alpha_s^* = \frac{\sum_{i=1}^{n} x_i^s}{n}; \qquad \mu_s^* = \frac{\sum_{i=1}^{n} (x_i - m_x^*)^s}{n}.$$

Статистичні початкові чи центральні моменти виражають необхідні числові характеристики статистичного розподілу випадкової величини. При достатньо великому n вони наближаються (збігаються за ймовірністю) до відповідних теоретичних аналогів і наближено дорівнюють їм.

Наведені розрахункові формули використовують, якщо первинний статистичний матеріал подано у формі простої статистичної сукупності. За умови великої кількості випробувань з метою оптимізації процесу розрахунків числових характеристик можна використати статистичний ряд

розподілу. У цьому випадку характеристики розраховують за відносними частотами p_i^* кожного з k інтервалів статистичного ряду. Значення, яких набуває випадкова величина в кожному інтервалі ряду можна вважати практично постійними. Тому представляти інтервал достатньо за наближеним значенням $\tilde{x}_i$, яке дорівнює середньому значенню величини в інтервалі, якщо воно відоме, або навіть середині інтервалу. Тоді статистичні початкові та центральні моменти і числові характеристики виражаються наближеними формулами такого вигляду:

$$m_x^* = \sum_{i=1}^{k} \tilde{x}_i p_i^* ; \qquad D_x^* = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^2 p_i^* ;$$

$$\alpha_s^* = \sum_{i=1}^{k} \tilde{x}_i^s p_i^* ; \qquad \mu_s^* = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^s p_i^* .$$

Правомірність використання розрахункових формул для числових характеристик статистичного розподілу обґрунтовується законом великих чисел, беручи до уваги теореми Чебишева і Бернуллі.

1. ВИЗНАЧЕННЯ ЗАКОНУ РОЗПОДІЛУ ВИПАДКОВОЇ ВЕЛИЧИНИ НА ОСНОВІ ДОСЛІДНИХ ДАНИХ

У практиці вирішення статистичних задач здебільшого обробляють набори первинних експериментальних даних, які одержані за відносно невеликою кількістю випробувань. Тоді потрібно рахуватися із тим, що статистичному розподілу можуть бути властиві нечіткі закономірності набутих результатів експерименту з певними ознаками випадковості. Тільки за умови великої кількості випробувань елементи випадковості згладжуються і випадкове явище повною мірою набуває властиву йому закономірність. Тому з метою визначення закону розподілу випадкової величини доводиться вирішувати задачу вибору для заданого статистичного ряду такої теоретичної кривої розподілу, яка б виражала лише суттєві ознаки первинного матеріалу і не була обтяжена елементами випадковості внаслідок недостатньої кількості експериментальних даних. У математичній статистиці таку задачу називають вирівнювання (згладжування) статистичного ряду. **Вирівнювання статистичного ряду розподілу** – це задача вибору теоретичної кривої розподілу, яка би з тих чи інших міркувань найкраще описувала заданий статистичний розподіл.

З огляду на окреслені умови, з математичної точки зору задача вирівнювання статистичного ряду розподілу не може мати однозначного строгого результату розв’язку. Тому прийнятність кінцевого результату може бути обґрунтована виключно конкретизацією найкращої відповідності вибраної теоретичної кривої статистичному розподілові з використанням якогось оціночного критерію і вираження такої відповідності у тій чи іншій числовій формі.

Принциповий вигляд теоретичної кривої вибирають завчасно з міркувань, які пов'язують зі змістом поставленої задачі. В одних випадках, задача формулюється з метою перевірки підпорядкованості заданого статистичного розподілу тому чи іншому конкретному теоретичному законові. У інших випадках, вибір кривої розподілу з відповідним їй аналітичним вираженням здійснюється візуально, виходячи із зовнішнього вигляду статистичного розподілу, який представлений графічно у формі гістограми. Наприклад, на схемі рис. 5 гістограмами зображено два різних статистичних розподіли і дві криві теоретичних законів розподілу. Одна крива відповідає нормальному розподілу з функцією щільності $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m_x)^2}{2\sigma^2}}$, друга – рівномірному розподілу з функцією $f(x)=\begin{cases}\frac{1}{\beta-\alpha} & \text{при} \quad \alpha<x<\beta \\ 0 & \text{при } x<\alpha \text{ або } x>\beta\end{cases}$. Навіть візуальне порівняння зображених гістограм із кривими теоретичних розподілів дає змогу зробити правильний вибір вирівнюючої кривої, як це показано на схемі.

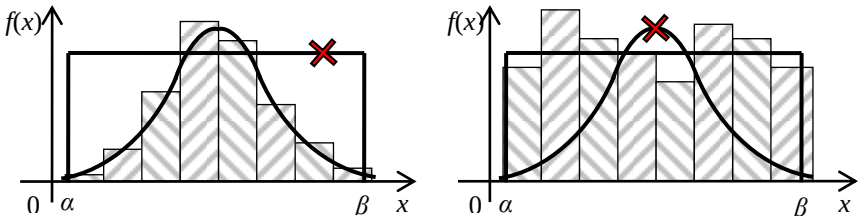


Рис. 5. Порівняльна схема до вибору вирівнюючої кривої для різних статистичних розподілів

Який би підхід не був використаний у формулюванні постановки задачі, вибрана крива лише у графічній формі представляє аналітичне вираження функції щільності теоретичного закону розподілу. Аналітичне вираження функції щільності завжди залежить від кількох параметрів – числових характеристик, які виражають ті чи інші ознаки теоретичного розподілу. Тому вирівнювання статистичного розподілу перетворюється у задачу визначення таких невідомих значень параметрів, з якими відповідність між статистичним і теоретичним розподілами буде найкращою. Наприклад, у разі вибору для вирівнювання статистичного розподілу кривої нормального закону потрібно вирішувати задачу визначення невідомих значень теоретичних параметрів m_x і σ – математичного сподівання і середнього квадратичного відхилення. Визначення параметрів потрібно здійснювати під

умовою, щоб крива теоретичного розподілу найкраще описувала заданий статистичний ряд розподілу. Разом з тим, необхідно дотримуватись властивостей теоретичної функції щільності розподілу.

Один з найбільш поширених методів вирішення такого завдання називається **метод моментів**. Суть методу полягає у тому, що невідомим параметрам теоретичного розподілу надають таких значень, щоб вони дорівнювали відповідним їм статистичним аналогам, які доступні елементарним обчисленням за даними статистичного ряду. Якщо аналітичне вираження функції $f(x)$ залежить, наприклад, від математичного сподівання m_X і середнього квадратичного відхилення σ, то їх прирівнюють відповідним статистичним аналогам m_X^* і σ^*, тобто накладають умову $m_X = m_X^*$ і $\sigma = \sigma^*$.

Вибором теоретичної кривої завершується вирішення завдання вирівнювання статистичного ряду розподілу. Проте одержаний результат не дає остаточної відповіді на предмет того, чи підпорядковується досліджувана випадкова величина вибраному теоретичному законові. Цей результат є лише гіпотезою, адже як би вдало не була підібрана крива, між нею і статистичним розподілом завжди неминучі розбіжності і їх походження та значущість досі не встановлено. Розбіжності можуть бути наслідком випадкових обставин, які пов'язані з недостатньою кількістю первинного експериментального матеріалу. Існуючі розбіжності можуть бути навіть настільки суттєвими, що зумовлені невдалим вибором теоретичної кривої розподілу. Тому для формулювання остаточної відповіді щодо ступеню узгодженості теоретичного і статистичного розподілів означена гіпотеза підлягає обов'язковій перевірці на предмет її правдоподібності. Підтвердити або, рівною мірою, спростувати істинність цієї гіпотези дають змогу критерії перевірки статистичних гіпотез.

Загальний зміст критеріїв перевірки статистичних гіпотез. Допустимо, на основі заданого експериментального матеріалу потрібно перевірити гіпотезу про підпорядкованість випадкової величини X теоретичному законові, який використовувавсь на стадії вирівнювання статистичного ряду. Якщо теоретичному законові властиві функції розподілу $F(x)$ і щільності розподілу $f(x)$, то дану гіпотезу формулюють як таку, що величина X має функцію розподілу $F(x)$. З метою підтвердження чи спростовування гіпотези використовують деяку випадкову величину U, яка виражає міру розбіжностей теоретичного і статистичного розподілів. Закон розподілу величини U залежить від закону розподілу величини X, над якою було проведено випробування, а також від кількості випробувань n. Якщо гіпотеза підтвердиться, то закон розподілу величини U буде повністю визначатись законом розподілу величини X і кількістю випробувань n.

Якщо міра розбіжностей U набула деякого значення u, то розкривається можливість з'ясувати, чи це значення є наслідком випадкових причин і гіпотеза може бути підтверджена як правдоподібна, чи значення u надто велике внаслідок суттєвих розбіжностей між теоретичним і статистичним розподілами, і тоді гіпотезу потрібно спростовувати як хибну. З цією метою визначають ймовірність події $U \geq u$ – це ймовірність того, що внаслідок випадкових причин, які пов'язані з недостатнім обсягом експериментального матеріалу, міра розбіжностей U виявиться не меншою, ніж її набуте в результаті випробувань значення u.

Залежно від значення, якого набуває ймовірність $p(U \geq u)$, на практиці приймають наступні рішення.

1. Якщо $p(U \geq u) < 0.1$, то гіпотезу про узгодженість статистичного і теоретичного розподілів потрібно спростувати як хибну. Тоді рекомендується повторити експеримент або переглянути гіпотезу за результатами вирівнювання статистичного ряду на інших умовах.
2. Якщо $p(U \geq u) \geq 0.1$, то гіпотезу слід визнати такою, що вона відповідає експериментальним даним і підтверджується як правдоподібна з наступними рівнями значущості:
   - якщо $0.1 \leq p(U \geq u) < 0.3$, то узгодженість задовільна;
   - якщо $0.3 \leq p(U \geq u) < 0.5$ – узгодженість добра;
   - якщо $p(U \geq u) \geq 0.5$ – узгодженість відмінна.

Залежно від вибору міри розбіжностей U використовують наступні критерії перевірки гіпотези про узгодженість статистичного і теоретичного розподілів.

Критерій χ^2 Пірсона – це один із найбільш строгих і найчастіше використовуваних критеріїв перевірки узгодженості.

Допустимо, для випадкової величини X за результатами n незалежних випробувань сформовано статистичний ряд розподілу (див. табл. 2) і за результатами вирівнювання ряду кривою деякого теоретичного закону з функціями $F(x)$ і $f(x)$ висунуто гіпотезу про підпорядкованість величини цьому законові.

Знаючи функцію розподілу $F(x)$, розраховують теоретичні ймовірності $p_1, p_2, ..., p_k$ попадання величини X в інтервали статистичного ряду. Міру розбіжностей U між теоретичним і статистичним розподілами виражають сумою квадратів відхилень $p_i^* - p_i$ між відносними частотами ряду p_i^* та відповідними ймовірностями p_i з урахуванням ваг відхилень вигляду $\frac{n}{p_i}$:

$$U = \sum_{i=1}^{k} \frac{n}{p_i} (p_i^* - p_i)^2 .$$

За такого вибору при великій кількості випробувань n закон розподілу величини U практично не залежить від функції розподілу $F(x)$ та значення n, а залежить виключно від кількості інтервалів статистичного ряду k. Закон розподілу величини U гранично наближається до так званого "розподілу χ^2". Беручи до уваги вираження відносних частот $p_i^* = \frac{m_i}{n}$, де m_i – це кількість значень величини X у кожному i-му інтервалі статистичного ряду, вираження міри розбіжностей U набуває вигляду

$$U = \chi^2 = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}.$$

Розподіл χ^2 залежить від параметра $r = k - s$, який називають число ступенів свободи. Тут s – кількість незалежних умов, якими обтяжене вирівнювання статистичного ряду розподілу. Умови s можуть бути різними за кількістю і за змістом. Це визначають відповідні умови вирівнювання статистичного ряду. Такими умовами можуть бути, наприклад, наступні:

1) $\sum_{i=1}^{k} p_i^* = 1$. Така умова має місце завжди, оскільки вона є обов'язковою на стадії побудови статистичного ряду розподілу і відповідає властивості функції щільності $\int_{-\infty}^{+\infty} f(x)dx = 1$;
2) $m_x = m_x^*$, якщо на стадії вирівнювання статистичного ряду теоретичний розподіл вибирають під умовою рівності теоретичного і статистичного значень математичного сподівання;
3) $\sigma = \sigma^*$, якщо на стадії вирівнювання статистичного ряду теоретичний розподіл вибирають під умовою рівності теоретичного і статистичного значень середнього квадратичного відхилення.

З метою практичного використання розподілу χ^2 складено спеціальні таблиці (див. додаток 1). За ними для обчисленого значення міри розбіжностей $u = \chi^2$ і числа ступенів свободи r можна визначити ймовірність $p(U \geq u)$ того, що величина $U = \chi^2$, яка розподілена за законом χ^2, перевищить набуте в результаті випробувань значення u. Ймовірність $p(U \geq u)$ дає змогу підтвердити чи спростувати гіпотезу про підпорядкованість випадкової величини X вибраному теоретичному законові.

Критерій Колмогорова А.Н. Мірою розбіжностей U між теоретичним і статистичним розподілами використовується максимальне значення модуля різниці D між статистичною функцією розподілу $F^*(x)$ і функцією

розподілу $F(x)$ теоретичного закону, який взято за основу вирівнювання статистичного ряду (рис. 6):

$$U = D = \max\left|F^*(x) - F(x)\right|.$$

Значення міри $u = D$ просто обчислюється – для цього достатньо визначити значення функцій $F^*(x)$ та $F(x)$ на границях інтервалів статистичного ряду розподілу і визначити їх максимальну розбіжність.

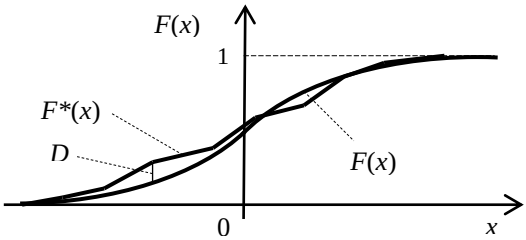


Рис. 6. Максимальна розбіжність *D* між значеннями статистичної *F**(*x*) і теоретичної *F*(*x*) функцій розподілу

Випадкова величини $U = D$ має досить простий закон розподілу. Якою б не була функція розподілу $F(x)$ неперервної випадкової величини *X*, при необмеженому зростанні кількості незалежних випробувань *n* ймовірність нерівності $D\sqrt{n} \geq \lambda$ прямує до границі

$$p(\lambda) = 1 - \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2}.$$

Для значень ймовірностей $p(\lambda)$, які розраховані за такою формулою, складено таблиці (див. додаток 2). $p(\lambda)$ – це ймовірності того, що за умови підпорядкованості випадкової величини *X* законові $F(x)$ внаслідок виключно випадкових причин максимальна розбіжність *D* між $F(x)$ та $F^*(x)$ буде не меншою, ніж фактично спостережена. Якщо ймовірність $p(\lambda)$ мала, то гіпотезу про узгодженість теоретичного і статистичного розподілів спростовують як хибну; при великих $p(\lambda)$ гіпотеза сумісна з дослідними даними.

Критерій Колмогорова досить простий, однак його рекомендують використовувати лише у тих випадках, коли теоретичний розподіл $F(x)$ відомий завчасно з тих чи інших міркувань, причому відомо не лише вигляд теоретичної функції розподілу $F(x)$, але й усі параметри, які входять до її складу. Такі випадки на практиці трапляються рідко. Здебільшого з певних

міркувань відомо лише загальний вигляд функції $F(x)$, а її числові параметри визначають на основі заданого статистичного матеріалу, як це роблять, наприклад, використовуючи метод моментів. Критерій Пірсона таку ваду усуває зменшенням числа ступенів свободи розподілу χ^2. Тому, застосовуючи критерій Колмогорова після вирівнювання статистичних рядів методом моментів, одержують завищене значення ймовірності $p(\lambda)$. З цієї причини можна прийняти правдоподібною таку гіпотезу, яка фактично погано узгоджується з дослідними даними. У таких випадках критерії Пірсона і Колмогорова забезпечують різні результати перевірки гіпотези.

Якщо статистичний розподіл випадкової величини вирівнюється з допомогою нормального закону і умова задачі дозволяє наближену перевірку відповідної гіпотези, то в такому випадку достатньо використати **наближений критерій**. Згідно властивостей нормального закону розподілу теоретичні значення його асиметрії S_k та ексцесу E_x дорівнюють нулю. Статистичні значення асиметрії S_k^* та ексцесу E_x^* розподілу випадкової величини обчислюються на підставі дослідних даних за формулами:

$$S_k^* = \frac{\mu_3^*}{\sigma^{*3}}; \qquad E_x^* = \frac{\mu_4^*}{\sigma^{*4}} - 3.$$

Тут μ_3^* і μ_4^* – статистичні центральні моменти третього і четвертого порядків; σ^* – статистичне середнє квадратичне відхилення. Числові значення S_k^* та E_x^* залежать від особливостей розподілу випадкової величини та об'єму дослідних даних і лише за умови $n \to \infty$ можуть прямувати до нуля. Гіпотезу про підпорядкованість випадкової величини X нормальному законові розподілу наближено можна вважати сумісною з дослідними даними, якщо відхилення від нуля статистичних значень коефіцієнта асиметрії S_k^* та ексцесу E_x^* не перевищують певних допустимих значень. Практично, гіпотезу наближено можна вважати підтвердженою, якщо одночасно виконуються умови нерівностей наступного вигляду:

$$\left.\begin{aligned} \left|S_k^*\right| &\le 3\sqrt{\frac{6(n-1)}{(n+1)(n+3)}} \\ \left|E_x^*\right| &\le 5\sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}} \end{aligned}\right\}.$$

Наближений критерій вигідно вирізняється серед інших простотою свого практичного використання, проте забезпечує лише наближену перевірку узгодженості розподілу випадкової величини з нормальним законом.

Хід роботи

Під час виконання роботи проводиться статистична обробка результатів випробувань, проведених над випадковою величиною, з метою визначення її закону розподілу. Вирішення завдання включає вирівнювання статистичного розподілу величини за допомогою теоретичного закону і перевірку гіпотези про підпорядкування величини цьому законові розподілу.

Завдання. Проведено n повторних незалежних вимірів кутів мережі тріангуляції, в результаті яких одержано числові значення кутів і відповідні їм похибки. Необхідно вирівняти статистичний ряд розподілу похибок вимірів за допомогою нормального закону розподілу і перевірити гіпотезу про підпорядкованість похибок цьому теоретичному законові.

Вхідні дані для виконання завдання зведено до таблиці додатку 7 (задано m_i – кількість похибок у k=8 інтервалах статистичного ряду; $i = 1,2,...,k$).

Приклад розв'язування завдання. Допустимо, значення похибок вимірів кутів розподілені в інтервалах статистичного ряду наступним чином:

I_i	-4; -3	-3; -2	-2; -1	-1; 0	0; 1	1; 2	2; 3	3; 4
m_i	11	24	78	139	148	85	32	10

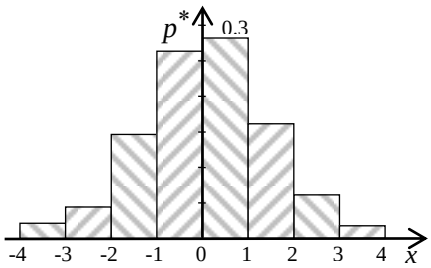
1. Вирівнювання статистичного розподілу похибок вимірів кутів за допомогою нормального закону розподілу.

1.1. Формування статистичного ряду розподілу. Для обчислення відносних частот похибок у кожному інтервалі використовуємо формулу $p_i^* = \dfrac{m_i}{n}$, беручи до уваги, що у поданому прикладі $n = \sum_{i=1}^{k} m_i = 527$:

I_i	-4; -3	-3; -2	-2; -1	-1; 0	0; 1	1; 2	2; 3	3; 4
p_i^*	0.021	0.045	0.148	0.264	0.281	0.161	0.061	0.019

Контроль обчислень: $\sum_{i=1}^{k} p_i^* = 1.000$.

1.2. Графічне зображення статистичного ряду у вигляді гістограми:



Взаємне розташування прямокутників гістограми за своїм зовнішнім виглядом подібне до форми кривої розподілу нормального закону. Це є

підставою виконати вирівнювання статистичного ряду за допомогою нормального закону розподілу. Вирівнювання реалізовуємо із застосуванням методу моментів. Метод передбачає, що теоретичні значення математичного сподівання m_x і середнього квадратичного відхилення σ, які входять до аналітичного вираження функції щільності $f(x)$ нормального закону, прирівнюють відповідним їм статистичним аналогам m_x^* і σ^*.

1.3. Обчислення статистичних значень математичного сподівання m_x^* і середнього квадратичного відхилення σ^* на основі статистичного ряду розподілу за формулами $m_x^* = \sum_{i=1}^{k} \tilde{x}_i p_i^*$, $D_x^* = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^2 p_i^*$, $\sigma^* = \sqrt{D_x^*}$. $\tilde{x}_i$ – середина кожного інтервалу статистичного ряду. Результати проміжних розрахунків відображено в таблиці:

I_i	-4; -3	-3; -2	-2; -1	-1; 0	0; 1	1; 2	2; 3	3; 4
$\tilde{x}_i$	-3.5	-2.5	-1.5	-0.5	0.5	1.5	2.5	3.5
p_i^*	0.021	0.045	0.148	0.264	0.281	0.161	0.061	0.019
$\tilde{x}_i p_i^*$	-0.074	-0.112	-0.222	-0.132	0.140	0.242	0.152	0.066
$(\tilde{x}_i - m_x^*)^2 p_i^*$	0.2661	0.2949	0.3602	0.0828	0.0544	0.3338	0.3632	0.2248

Остаточно одержимо: $m_x^* = 0.060$; $D_x^* = 1.9802$; $\sigma^* = 1.407$.

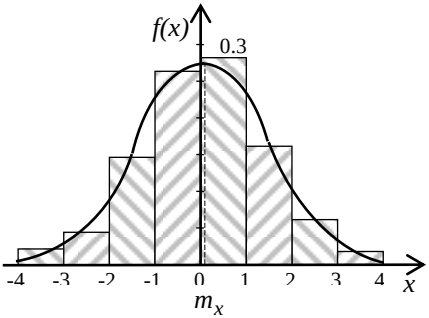
1.4. Побудова кривої нормального закону розподілу. Скористаємось аналітичним вираженням функції щільності нормального закону вигляду $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m_x)^2}{2\sigma^2}}$ і обчислимо значення функції $f(\tilde{x}_i)$ для середини інтервалів $\tilde{x}_i$, беручи до уваги, що $m_x = m_x^*$ та $\sigma = \sigma^*$. З цією метою, завчасно обчисливши нормовані значення $t_i = \frac{\tilde{x}_i - m_x}{\sigma}$, використаємо таблиці функції $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ (додаток 3). Остаточно $f(\tilde{x}_i) = \frac{f(t_i)}{\sigma}$. Результати обчислень відображено в таблиці:

$\tilde{x}_i$	-3.5	-2.5	-1.5	-0.5	0.5	1.5	2.5	3.5
t_i	-2.53	-1.82	-1.11	-0.40	0.31	1.02	1.73	2.44
$f(t_i)$	0.0163	0.0761	0.2155	0.3683	0.3802	0.2371	0.0893	0.0203
$f(\tilde{x}_i)$	0.012	0.054	0.153	0.262	0.270	0.168	0.063	0.014

Примітка: якщо довжини інтервалів не дорівнюють одиниці, то для визначення кінцевих значень функції щільності обчислені в останньому рядку таблиці значення $f(\tilde{x}_i)$ потрібно множити на довжини інтервалів.

Контроль обчислень: $\sum_{i=1}^{k} f(\tilde{x}_i) = 0.996 \approx 1.$

За значеннями $f(\tilde{x}_i)$ на одному малюнку з гістограмою статистичного розподілу будуємо вирівнюючу криву нормального закону розподілу:



Одержане зображення показує, що крива нормального закону розподілу добре описує гістограму статистичного розподілу. На цій підставі можна висунути гіпотезу, що похибки вимірів кутів підпорядковуються нормальному законові розподілу.

2. Перевірка правдоподібності гіпотези про підпорядкованість похибок вимірів кутів нормальному законові розподілу

2.1. Перевірка гіпотези за критерієм χ^2 Пірсона.

Теоретичні ймовірності p_i попадання нормально розподілених похибок вимірів у інтервали статистичного ряду розподілу обчислюємо за формулою $p_i = \Phi^*(t_{i_{пр}}) - \Phi^*(t_{i_л}) = \Phi^*(\frac{x_{i_{пр}} - m_x^*}{\sigma^*}) - \Phi^*(\frac{x_{i_л} - m_x^*}{\sigma^*})$. Тут позначено: $t_{i_{пр}}$ та $t_{i_л}$ – нормовані аргументи на правій $x_{i_{пр}}$ та лівій $x_{i_л}$ границях інтервалів. Значення нормальної функції розподілу $\Phi^*(t_{i_{пр}})$ та $\Phi^*(t_{i_л})$ визначаємо з таблиці додатку 4. Результати обчислень відображено в таблиці:

I_i	-4; -3	-3; -2	-2; -1	-1; 0	0; 1	1; 2	2; 3	3; 4	
$x_{i_л}; x_{i_{пр}}$	-4	-3	-2	-1	0	1	2	3	4
$t_{i_л}; t_{i_{пр}}$	-2.886	-2.175	-1.464	-0.753	-0.043	0.668	1.379	2.090	2.800
$\Phi^*(t_i)$	0.0020	0.0149	0.0716	0.2257	0.4828	0.7480	0.9161	0.9816	0.9974
p_i	0.013	0.057	0.154	0.257	0.265	0.168	0.066	0.016	

Контроль обчислень: $\sum_{i=1}^{k} p_i = 0.996 \approx 1$.

Значення u міри розбіжностей U між статистичним і теоретичним розподілами обчислюємо за формулою $u = \chi^2 = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}$:

I_i	-4;-3	-3;-2	-2;-1	-1;0	0; 1	1; 2	2; 3	3; 4
p_i	0.013	0.057	0.154	0.257	0.265	0.168	0.066	0.016
m_i	11	24	78	139	148	85	32	10
np_i	6.851	30.039	81.158	135.439	139.655	88.536	34.782	8.432
$\frac{(m_i - np_i)^2}{np_i}$	2.513	1.214	0.123	0.094	0.499	0.141	0.222	0.292

Остаточно одержимо: $u = \chi^2 = 5.098$.

У задачах вирівнювання статистичних рядів за допомогою нормального закону методом моментів кількість накладених незалежних умов $s = 3$. На цій основі кількість ступенів свободи розподілу χ^2 складає $r = k - s = 5$. Тепер за значеннями $\chi^2 = 5.098$ та $r = 5$ з таблиці додатку 1 визначаємо ймовірність $p(\chi^2) = 0.42$, що вкладається у межах $0.3 \le p(\chi^2) < 0.5$.

Висновок. Ймовірність $p(\chi^2) = 0.42$ того, що міра розбіжностей χ^2 не менша свого обчисленого значення, є достатньо значущою. Цим посвідчується факт, що існуючі розбіжності між теоретичним і статистичним розподілами викликані випадковими причинами, пов'язаними з обмеженою кількістю випробувань. При $0.3 \le p(\chi^2) < 0.5$ узгодженість розподілів добра. Отже, гіпотезу про підпорядкованість похибок вимірів кутів нормальному законові можна підтвердити як таку, що не суперечить дослідним даним.

2.2.Перевірка гіпотези за критерієм А.Н.Колмогорова.

Значення статистичної функції розподілу похибок вимірів $F^*(x)$ обчислюємо за даними статистичного ряду, обираючи значеннями змінної x границі інтервалів, з використанням формули $F^*(x) = \sum_{x_i < x} p_i^*$:

$F^*(-4) = 0;$ $F^*(-3) = p_1^* = 0.021;$ $F^*(-2) = p_1^* + p_2^* = 0.066;$

$F^*(-1) = p_1^* + p_2^* + p_3^* = 0.214;$ $F^*(0) = p_1^* + \ldots + p_4^* = 0.478;$

$F^*(1) = p_1^* + \ldots + p_5^* = 0.759;$ $F^*(2) = p_1^* + \ldots + p_6^* = 0.920;$

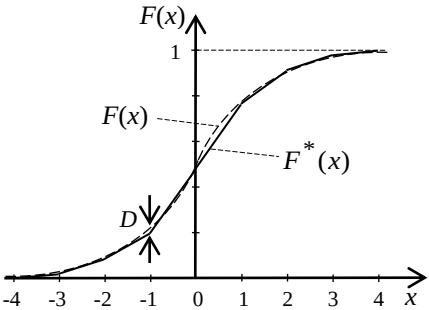
$F^*(3) = p_1^* + \ldots + p_7^* = 0.981;$ $F^*(4) = p_1^* + \ldots + p_8^* = 1.000.$

Значення теоретичної функції розподілу $F(x)$ для нормального закону визначаються за посередництва нормальної функції розподілу $\Phi^*(t)$ за

таблицями додатку 4, як це було реалізовано у п.2.1: $F(x_i) = \Phi^*(t_i)$; $t_i = \dfrac{x_i - m_x^*}{\sigma^*}$ – це нормовані значення змінної x_i, якою почергово обираємо границі інтервалів статистичного ряду.

Результати обчислення значень статистичної $F^*(x)$ і теоретичної $F(x)$ функцій розподілу, а також їх абсолютних розбіжностей $\left|F^*(x_i) - F(x_i)\right|$, поміщено до таблиці та відображено на графіку:

x_i	-4	-3	-2	-1	0	1	2	3	4
$F(x)$	0.002	0.015	0.072	0.226	0.483	0.748	0.916	0.982	0.997
$F^*(x)$	0	0.021	0.066	0.214	0.478	0.759	0.920	0.981	1.000
$\left|F^*(x_i) - F(x_i)\right|$	0.002	0.006	0.006	0.012	0.005	0.011	0.004	0.001	0.003



Одержані результати показують, що значення міри розбіжностей $u=D$, яка виражається максимальною різницею значень статистичної та теоретичної функцій розподілу, відповідає значенню змінної $x = -1$: $D = \max\left|F^*(x_i) - F(x_i)\right| = 0.012$. Обчислюємо величину $\lambda = D\sqrt{n} = 0.28$ і за таблицею додатку 2 визначаємо ймовірність $p(\lambda)$ того, що за умови підпорядкованості похибок вимірів кутів законові розподілу $F(x)$ внаслідок виключно випадкових причин максимальна розбіжність D між $F(x)$ та $F*(x)$ буде не меншою, ніж фактично обчислена: $p(\lambda) = 1.000$.

Висновок. Подію з ймовірністю $p(\lambda) = 1.000$ вважають достовірною. Отже, узгодженість статистичного і теоретичного розподілів відмінна і гіпотезу про підпорядкованість похибок вимірів кутів нормальному законові розподілу можна вважати сумісною з дослідними даними.

У даній задачі має місце випадок, коли параметри теоретичного закону розподілу не відомі. У ході розв'язку теоретичні значення математичного сподівання та середнього квадратичного відхилення прирівнювались значенням їх статистичних аналогів. З цієї причини значення ймовірності $p(\lambda) = 1.000$ суттєво завищене порівняно з відповідною ймовірністю $p(\chi^2) = 0.42$ критерію Пірсона. Критерій Колмогорова міг би забезпечити значення ймовірності такого ж порядку, як і критерій Пірсона виключно за умови, якби теоретичний розподіл був відомий разом з його параметрами.

2.3. Перевірка гіпотези за наближеним критерієм.

Статистичні значення коефіцієнта асиметрії S_k^* та ексцесу E_x^* обчислюємо за формулами $S_k^* = \dfrac{\mu_3^*}{\sigma^{*3}}$ і $E_x^* = \dfrac{\mu_4^*}{\sigma^{*4}} - 3$, де $\mu_3^* = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^3 p_i^*$ і $\mu_4^* = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^4 p_i^*$ – статистичні центральні моменти третього і четвертого порядків. Результати проміжних обчислень:

$\tilde{x}_i$	-3.5	-2.5	-1.5	-0.5	0.5	1.5	2.5	3.5
p_i^*	0.021	0.045	0.148	0.264	0.281	0.161	0.061	0.019
$(\tilde{x}_i - m_x^*)^3 p_i^*$	-0.947	-0.755	-0.562	-0.046	0.024	0.481	0.886	0.773
$(\tilde{x}_i - m_x^*)^4 p_i^*$	3.373	1.933	0.876	0.026	0.010	0.692	2.162	2.661

Остаточно одержимо: $\mu_3^* = -0.146$; $\mu_4^* = 11.733$; $S_k^* = -0.052$; $E_x^* = -0.006$.

Статистичні значення асиметрії та ексцесу не дорівнюють нулю, як це властиво нормальному законові розподілу. Асиметрія статистичного розподілу від'ємна, тобто крива розподілу зміщена вліво порівняно з кривою нормального закону. Ексцес статистичного розподілу констатує, що крива цього розподілу є більш плоско вершинною порівняно з кривою нормального закону. Допустимі відхилення від нуля коефіцієнта асиметрії та ексцесу перевіряємо за умовами нерівностей

$$\left.\begin{aligned} \left|S_k^*\right| &\le 3\sqrt{\frac{6(n-1)}{(n+1)(n+3)}} \\ \left|E_x^*\right| &\le 5\sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}} \end{aligned}\right\}: \qquad \left.\begin{aligned} \left|S_k^*\right| &= 0.052 < 0.315 \\ \left|E_x^*\right| &= 0.006 < 1.049 \end{aligned}\right\}.$$

Висновок. Умови обох нерівностей виконуються, відхилення від нуля коефіцієнта асиметрії та ексцесу не перевищують допустимих значень. На цій підставі гіпотеза про підпорядкованість похибок вимірів кутів нормальному законові наближено може бути прийнята як правдоподібна.

Питання для самоконтролю

1. Що називають простою статистичною сукупність?
2. Що називають статистичним рядом розподілу?
3. Що таке гістограма?
4. Що таке статистична функція розподілу і як її побудувати?
5. Які числові характеристики має статистичний розподіл?
6. Поясніть зміст завдання вирівнювання статистичного ряду розподілу. Як розв'язується це завдання?
7. Які особливості вирівнювання статистичного ряду розподілу з допомогою нормального закону?
8. Поясніть зміст критеріїв перевірки статистичних гіпотез.
9. Поясніть зміст критерію Пірсона. Який порядок його практичного застосування?
10. Поясніть зміст критерію Колмогорова. Який порядок його практичного застосування?
11. Як можна наближено перевірити підпорядкованість випадкової величини нормальному законові розподілу на основі дослідних даних?

2. ВИЗНАЧЕННЯ ТОЧНОСТІ ТА НАДІЙНОСТІ ЧИСЛОВИХ ОЦІНОК ПАРАМЕТРІВ РОЗПОДІЛУ В УМОВАХ ОБМЕЖЕНОЇ КІЛЬКОСТІ ВИПРОБУВАНЬ

Визначення закону розподілу випадкової величини можливе за умови наявності достатньо об'ємного первинного статистичного матеріалу – порядку хоча б кількох перших сотень результатів випробувань. На практиці часто такий матеріал обмежений кількома десятками випробувань. Обмежений обсяг первинного матеріалу унеможливлює вирішення задачі визначення закону розподілу випадкової величини. Проте його можна використати для формулювання певних відомостей про величину, зокрема, в частині хоча б наближеного описування окремих ознак її розподілу тими чи іншими числовими характеристиками. Наближені значення параметрів розподілу випадкової величини, які розраховані в умовах обмеженої кількості випробувань, містять значні елементи випадковості. Це зумовлює певні особливості вирішення окресленого завдання, недотримання яких має наслідком формальні результати статистичної обробки первинного матеріалу.

Наближене значення невідомого параметра a розподілу випадкової величини, яке розраховане в умовах обмеженої кількості випробувань, називається оцінка $\tilde{a}$ параметра a; оскільки $\tilde{a}$ розкривається окремим числом, то її називають **числова оцінка** $\tilde{a}$ **параметра** a.

Допустимо, проведено обмежену кількість n незалежних випробувань, у кожному з яких випадкова величина X набула значень $x_1, x_2, ..., x_n$. Щоб за цими результатами визначити найкраще наближення оцінки $\tilde{a}$ для параметра a, вона повинна задовольняти наступним вимогам:

1) оцінка $\tilde{a}$ повинна бути найбільш надійною: за умови збільшення кількості випробувань n оцінка $\tilde{a}$ повинна наближатись (прямувати за ймовірністю) до самого параметра a, тобто $\tilde{a} \to a$ при $n \to \infty$;
2) оцінка $\tilde{a}$ повинна бути не зміщеною, для чого вона не повинна бути обтяжена систематичними похибками, а її математичне сподівання повинно дорівнювати самому параметрові, тобто $M[\tilde{a}] = a$;
3) оцінка $\tilde{a}$ повинна бути ефективною, для чого її дисперсія повинна бути мінімальною: $D[\tilde{a}] = \min$.

Виходячи з таких вимог, для основних параметрів статистичного розподілу випадкових величин використовують наступні числові оцінки.

Найбільш надійна, не зміщена та ефективна **числова оцінка** $\tilde{m}_x$ **математичного сподівання** m_x виражається середнім статистичним m_x^*:

$$\tilde{m}_x = m_x^* = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Формула статистичної дисперсії D_x^* у чистому вигляді непридатна для вираження відповідної числової оцінки $\tilde{D}_x$. Величина D_x^* не задовольняє вимогу не зміщеної оцінки дисперсії D_x, оскільки обтяжена систематичною похибкою, яка зумовлює заниження дисперсії на величину $\frac{n-1}{n}$. Ця похибка враховується введенням поправки $\frac{n}{n-1}$, тому найбільш надійна, не зміщена та ефективна **числова оцінка** $\tilde{D}_x$ **дисперсії** D_x виражається величиною

$$\tilde{D}_x = \frac{n}{n-1} D_x^* = \frac{\sum_{i=1}^{n} (x_i - \tilde{m}_x)^2}{n-1}.$$

У більшості статистичних задач потрібно не лише визначити для параметра a його числову оцінку $\tilde{a}$, але й схарактеризувати точність та надійність цього значення. Така потреба виникає тому, що постає питання визначення похибки заміни параметра a його числовою оцінкою $\tilde{a}$, а також оцінювання ступеня впевненості, що ця похибка не буде перевищувати задані межі. Зазначене особливо актуальне за умови малої кількості випробувань, адже тоді оцінка $\tilde{a}$ значною мірою випадкова і наближена заміна a на $\tilde{a}$ може спричинити хибну інтерпретацію параметра. Критеріями точності та надійності числової оцінки $\tilde{a}$ є довірчий інтервал та довірча ймовірність.

Надійність оцінки $\tilde{a}$ параметра a виражає **довірча ймовірність** β. Довірча ймовірність β визначає ступінь довіри до похибки ε, яка виникає внаслідок заміни a на $\tilde{a}$: $\beta = p(|\tilde{a} - a| < \varepsilon)$ або $\beta = p(\tilde{a} - \varepsilon < a < \tilde{a} + \varepsilon)$.

Отже, з ймовірністю β невідоме значення параметра a потрапить у інтервал $I_\beta = (\tilde{a} - \varepsilon; \tilde{a} + \varepsilon)$, а діапазон практично можливих значень похибки заміни a на $\tilde{a}$ буде $\pm\varepsilon$. За своєю сутністю β – це ймовірність практично достовірної події. Великі за абсолютною величиною похибки ε можливі тільки з малою ймовірністю $\alpha = 1 - \beta$. Виходячи з таких міркувань, довірчу ймовірність встановлюють апріорі на рівні 0.9 ÷ 0.99. Інтервал I_β називають довірчий інтервал, а границі $a_1 = \tilde{a} - \varepsilon$ та $a_2 = \tilde{a} + \varepsilon$ інтервалу I_β – довірчі границі. **Довірчий інтервал** I_β виражає точність числової оцінки $\tilde{a}$ (рис. 7).

Довірчий інтервал I_β визначають його центр $\tilde{a}$ і похибка ε заміни a на $\tilde{a}$. Значення $\tilde{a}$ і ε обчислюються як функції незалежних випадкових результатів випробувань $x_1, x_2, ..., x_n$. Внаслідок цього інтервал I_β і його положення на числовій осі також є випадковими. Тому β – це ймовірність того, що випадковий інтервал I_β покриває невідому точку a.

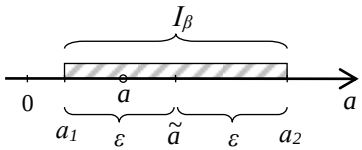


Рис. 7. Графічна інтерпретація
точності заміни параметра a числовою оцінкою $\tilde{a}$

Довірчий інтервал I_β слід розуміти як інтервал можливих значень невідомого параметра a, які сумісні з дослідними даними і які відповідають їм. Таке трактування довірчого інтервалу має наступне обґрунтування: якщо подію з ймовірністю $\alpha = 1 - \beta$ вважати практично неможливою, то ті значення параметра a, для яких $|\tilde{a} - a| > \varepsilon$, є несумісні з дослідними даними і такими, що не відповідають їм, тоді як значення параметра a, для яких $|\tilde{a} - a| < \varepsilon$, є сумісні з дослідними даними і відповідають їм.

Похибка ε заміни a на $\tilde{a}$ значною мірою залежна від того, чи відомо апріорі закон розподілу випадкової величини. Це зумовлює неоднозначність підходів до визначення ε та I_β, але врегульовується використанням точних чи наближених методів побудови довірчих інтервалів.

Точні методи побудови довірчих інтервалів для математичного сподівання та дисперсії мають наступний зміст. Будь-який довірчий інтервал для невідомого параметра a визначають під умовою ймовірності виконання

нерівності, яка містить його числову оцінку $\tilde{a}$. Закон розподілу випадкової числової оцінки $\tilde{a}$ залежить від самих невідомих параметрів розподілу випадкової величини *X*. На основі цієї нерівності здійснюється перехід від випадкової величини $\tilde{a}$ до деякої функції спостережених значень величини *X*. Закон розподілу такої функції не залежить від невідомих параметрів, а залежить тільки від кількості випробувань *n* і від виду закону розподілу величини *X*. Такого виду функції розглядаються лише для нормального розподілу величини *X*. Тому точні методи побудови довірчих інтервалів для невідомих параметрів *a* розподілу випадкової величини *X* використовують виключно у тих випадках, коли заздалегідь відомо, що ця величина розподілена за нормальним законом.

Доведено, що за умови нормального розподілу величини *X* випадкова величина $T = \sqrt{n}\,\dfrac{\tilde{m}_X - m_X}{\sqrt{\tilde{D}_X}}$ підпорядковується законові розподілу Стьюдента з $n-1$ ступенями свободи. На такій основі, задаючись довірчою ймовірністю β, будують **довірчий інтервал для математичного сподівання** m_X:

$$I_\beta = (\tilde{m}_X - \varepsilon_\beta; \tilde{m}_X + \varepsilon_\beta).$$

Похибку ε_β заміни m_X на $\tilde{m}_X$ під умовою $p(|\tilde{m}_X - m_X| < \varepsilon_\beta) = \beta$ виражає формула $\varepsilon_\beta = t_\beta \sqrt{\dfrac{\tilde{D}_X}{n}}$. Параметр t_β визначається за ймовірністю β з інтеграла $2\int\limits_0^{t_\beta} S_{n-1}(t)dt = \beta$ під умовою $p(|T| < t_\beta) = \beta$. Тут $S_{n-1}(t)$ – це щільність закону розподілу Стьюдента з $n-1$ ступенями свободи. Для зазначеного інтеграла складено спеціальну таблицю (див. додаток 5). Користуючись таблицею, за значенням довірчої ймовірності β і числом ступенів свободи $n-1$ визначають потрібний параметр t_β.

Доведено, що за умови нормального розподілу величини *X* випадкова величина $V = \dfrac{(n-1)\tilde{D}_X}{D_X}$ має розподіл χ^2 Пірсона з $r = n-1$ ступенями свободи. Випадкова величина $\tilde{D}_X$ виражається через величину V залежністю $\tilde{D}_X = V\dfrac{D_X}{n-1}$. На такій основі, задаючись довірчою ймовірністю β будують **довірчий інтервал для дисперсії** D_X:

$$I_\beta = (D_1; D_2).$$

Довірчий інтервал I_β з границями D_1 і D_2, який покриває невідоме значення дисперсії D_X з ймовірністю $\beta = p(D_1 < D_X < D_2)$, покриває його

тоді і тільки тоді, коли величина V потрапляє у відповідний їй інтервал i_β. Отже, інтервал I_β для дисперсії можна знайти за інтервалом i_β для величини V. Інтервал $i_\beta = (\chi_1^2; \chi_2^2)$ з довірчими границями χ_1^2 та χ_2^2, в який потрапляють значення величини V із заданою ймовірністю β, можна визначити за законом розподілу χ^2 цієї величини. Щоб побудувати інтервал i_β, використовують таблицю розподілу χ^2 (див. додаток 1). В ній поміщено значення χ^2 для величини V, яка має розподіл χ^2 з $r = n-1$ ступенями свободи такі, що $p(V > \chi^2) = p$. Якщо в таблиці зафіксувати значення $r = n-1$, то можна знайти для нього два значення χ^2: одне – χ_1^2 за ймовірністю $p_1 = \frac{\alpha}{2}$; друге – χ_2^2 за ймовірністю $p_2 = 1 - \frac{\alpha}{2}$. Тут $\alpha = 1 - \beta$ – це ймовірність попадання величини V поза межі інтервалу i_β. Потрібні границі D_1 та D_2 довірчого інтервалу I_β виражають формули

$$D_1 = \frac{(n-1)\tilde{D}_x}{\chi_1^2}, \qquad D_2 = \frac{(n-1)\tilde{D}_x}{\chi_2^2}.$$

Наближені методи побудови довірчих інтервалів для математичного сподівання та дисперсії використовують у випадках, коли попереднє встановлення закону розподілу випадкової величини не є обов'язковою умовою, а вид закону розподілу не позначається суттєво на визначенні точності та надійності числових оцінок цих параметрів.

Числова оцінка $\tilde{m}_x$ є сумою *n* незалежних однаково розподілених значень x_i випадкової величини *X*. Відтак, за умовою центральної граничної теореми, $\tilde{m}_x$ розподілена за нормальним законом. Навіть при *n* порядку 10÷20 закон розподілу суми наближається до нормального. На такій основі можна визначити значення похибки ε_β при заданій ймовірності β, для якої $p(\tilde{m}_x - \varepsilon_\beta < m_x < \tilde{m}_x + \varepsilon_\beta) = \beta$. Ймовірність β виражається через нормальну функцію розподілу $\beta = 2\Phi^*\left(\frac{\varepsilon_\beta}{\sigma_{\tilde{m}}}\right) - 1$, звідки $\varepsilon_\beta = \sigma_{\tilde{m}} \arg\Phi^*\left(\frac{1+\beta}{2}\right)$. Тут $\arg\Phi^*\left(\frac{1+\beta}{2}\right)$ – це функція, обернена до $\Phi^*\left(\frac{1+\beta}{2}\right)$, тобто таке значення аргументу, при якому нормальна функція дорівнює $\frac{1+\beta}{2}$; $\sigma_{\tilde{m}} = \sqrt{\frac{\tilde{D}_x}{n}}$ –

середнє квадратичне відхилення оцінки $\tilde{m}_x$. Отже, при заданій довірчій ймовірності β **довірчий інтервал для математичного сподівання** m_x визначають його числова оцінка $\tilde{m}_x$ та похибка ε_β:

$$I_\beta = (\tilde{m}_x - \varepsilon_\beta; \tilde{m}_x + \varepsilon_\beta).$$

Числова оцінка $\tilde{D}_x$ є сумою n випадкових величин $\frac{(x_i - \tilde{m}_x)^2}{n-1}$. Ці величини залежні, оскільки кожна з них включає числову оцінку $\tilde{m}_x$, але закон розподілу такої суми також гранично наближається до нормального. Практично така гіпотеза може бути прийнятною навіть при n порядку 20÷30. На такій основі, **довірчий інтервал для дисперсії** D_x при заданій довірчій ймовірності $\beta = p(\tilde{D}_x - \varepsilon_\beta < D_x < \tilde{D}_x + \varepsilon_\beta)$ визначають його числова оцінка $\tilde{D}_x$ та похибка ε_β:

$$I_\beta = (\tilde{D}_x - \varepsilon_\beta; \tilde{D}_x + \varepsilon_\beta),$$

де $\varepsilon_\beta = \sigma_{\tilde{D}} \arg \Phi^*\left(\frac{1+\beta}{2}\right)$; $\sigma_{\tilde{D}} = \tilde{D}_x \sqrt{\frac{2}{n-1}}$ – наближене значення середнього квадратичного відхилення оцінки $\tilde{D}_x$

Хід роботи

Під час виконання роботи обчислюються числові оцінки математичного сподівання та дисперсії, визначаються їх точність і надійність точними та наближеними методами. На практиці при побудові довірчих інтервалів залежно від умов завдання користуються або точними, або наближеними методами. Використання тут разом точних і наближених методів переслідує мету засвоєння правил і порядку практичного використання різних методів.

Завдання. Проведено n=20 результатів вимірів випадкової величини X, за результатами яких розраховано статистичні значення математичного сподівання m_x^* та дисперсії D_x^*. Визначити відповідні цим параметрам числові оцінки $\tilde{m}_x$ та $\tilde{D}_x$ і побудувати довірчі інтервали точними та наближеними методами при довірчій ймовірності β =0.9.

Вхідні дані: статистичні значення математичного сподівання m_x^* та дисперсії D_x^* використати з результатів розв'язання завдання 1.

Приклад розв'язування завдання. Допустимо, m_x^*=0.060, D_x^*=1.9802.

1.Обчислення числових оцінок $\tilde{m}_x$ та $\tilde{D}_x$.

Числова оцінка математичного сподівання $\tilde{m}_x = m_x^* = 0.060$. Числова оцінка дисперсії $\tilde{D}_x = \frac{n}{n-1} D_x^* = 2.0844$.

2. Побудова довірчих інтервалів точними методами.

2.1. Довірчий інтервал для математичного сподівання $I_\beta = (\tilde{m}_x - \varepsilon_\beta; \tilde{m}_x + \varepsilon_\beta)$ будуємо з використанням формули $\varepsilon_\beta = t_\beta \sqrt{\frac{\tilde{D}_x}{n}}$. Параметр t_β визначаємо з таблиці розподілу Стьюдента (додаток 5) за довірчою ймовірністю β =0.9 і числом ступенів свободи n–1=19: t_β =1.729. Отже, похибка $\varepsilon_\beta = 0.558$ і довірчий інтервал I_β = (–0.498; 0.618).

2.2. Довірчий інтервал для дисперсії $I_\beta = (D_1; D_2)$ будуємо з використанням формул $D_1 = \frac{(n-1)\tilde{D}_x}{\chi_1^2}$ і $D_2 = \frac{(n-1)\tilde{D}_x}{\chi_2^2}$. Значення величин χ_1^2 та χ_2^2 визначаємо з таблиці розподілу χ^2 (додаток 1): за числом ступенів свободи $r = n - 1 = 19$ та ймовірністю $p_1 = \frac{1-\beta}{2} = 0.05$ визначаємо значення χ_1^2 =30.1, за ймовірністю $p_2 = 1 - \frac{1-\beta}{2} = 0.95$ – значення χ_2^2 =10.11. Отже, довірчий інтервал для дисперсії I_β = (1.3157; 3.9173).

3. Побудова довірчих інтервалів наближеними методами.

3.1. Довірчий інтервал для математичного сподівання $I_\beta = (\tilde{m}_x - \varepsilon_\beta; \tilde{m}_x + \varepsilon_\beta)$ будуємо з використанням формули $\varepsilon_\beta = \sigma_{\tilde{m}} \arg \Phi^*(\frac{1+\beta}{2})$. $\sigma_{\tilde{m}} = \sqrt{\frac{\tilde{D}_x}{n}} = 0.3228$. Значення аргументу, при якому нормальна функція розподілу дорівнює $\frac{1+\beta}{2} = 0.95$, визначаємо з таблиці додатку 4: $\arg \Phi^*(0.95) = 1.645$. Отже, похибка $\varepsilon_\beta = 0.531$ і довірчий інтервал I_β = (–0.471; 0.591).

3.2. Довірчий інтервал для дисперсії $I_\beta = (\tilde{D}_x - \varepsilon_\beta; \tilde{D}_x + \varepsilon_\beta)$ будуємо з використанням формули $\varepsilon_\beta = \sigma_{\tilde{D}} \arg \Phi^*(\frac{1+\beta}{2})$: $\sigma_{\tilde{D}} = \tilde{D}_x \sqrt{\frac{2}{n-1}} = 0.6763$; $\arg \Phi^*(0.95) = 1.645$. Отже, похибка $\varepsilon_\beta = 1.1125$ і довірчий інтервал I_β =(0.9719; 3.1969).

Питання для самоконтролю

1. Що таке числова оцінка невідомого параметра розподілу випадкової величини?
2. Яким вимогам повинна задовольняти числова оцінка невідомого параметра розподілу?

3. Як визначити числові оцінки математичного сподівання та дисперсії?
4. Що називають довірчою ймовірністю й довірчим інтервалом?
5. Який зміст завдання оцінки точності та надійності невідомого параметра розподілу?
6. Як оцінити точність та надійність числової оцінки математичного сподівання точним методом?
7. Як оцінити точність та надійність числової оцінки дисперсії точним методом?
8. Як оцінити точність та надійність числової оцінки математичного сподівання наближеним методом?
9. Як оцінити точність та надійність числової оцінки дисперсії наближеним методом?
10. В яких випадках для оцінки точності та надійності невідомих параметрів розподілу випадкової величини використовують приблизні методи? В яких випадках використовують точні методи?

3. ВИЗНАЧЕННЯ ТІСНОТИ ТА ФОРМИ КОРЕЛЯЦІЙНОГО ЗВ'ЯЗКУ В СИСТЕМІ ДВОХ ВИПАДКОВИХ ВЕЛИЧИН

Математична статистика розв'язує задачі, в яких результат випробування описується не однією, а двома чи більше випадковими величинами. У таких випадках ці величини утворюють систему. Систему випадкових величин $X, Y, ..., W$ позначують $(X, Y, ..., W)$. У даному завданні розглядаємо випадок найпростішої системи (X, Y) двох випадкових величин X, Y.

Властивості системи не вичерпуються властивостями окремих величин, які її утворюють. Вони включають також взаємні зв'язки (залежності) між випадковими величинами. Систему можуть утворювати залежні та незалежні випадкові величини. Залежність чи незалежність X та Y завжди взаємні: величини X та Y незалежні, якщо закон розподілу кожної з них не залежить від того, якого значення набула друга величина; величини X та Y залежні, якщо закон розподілу кожної з них залежить від того, якого значення набула друга величина.

Поняття залежності випадкових величин в теорії ймовірностей має особливий зміст. Залежність випадкових величин X та Y – це не строга функціональна залежність з математичної точки зору, коли за значенням однієї величини можна прямо визначити значення другої. Теорія ймовірностей розглядає випадок імовірнісної залежності, коли за значенням однієї величини можна визначити закон розподілу другої величини, тобто тенденцію зміни значень другої величини, але не її точне значення. Функціональна залежність величин X та Y – це граничний випадок імовірнісної залежності, коли її тіснота максимальна. Встановлення та вираження саме такої властивості системи випадкових величин являє собою актуальне та важливе з практичної точки зору завдання математичної

статистики. Система дій зі встановлення зв'язку в системі випадкових величин з метою вираження його тісноти і форми, називається **кореляційний аналіз**. Кореляційний аналіз включає розрахунок низки числових характеристик як окремих величин системи, так і системи загалом. Усі характеристики виражаються через початкові або центральні моменти.

Початковий момент $\alpha_{k,s}$ порядку k,s системи (X,Y) – це математичне сподівання добутку X^k та Y^s: $\alpha_{k,s} = M[X^k Y^s]$. Центральний момент $\mu_{k,s}$ порядку k,s системи (X,Y) – це математичне сподівання добутку k-ої та s-ої степені відповідних центрованих величин $\overset{\circ}{X} = X - m_x$ та $\overset{\circ}{Y} = Y - m_y$: $\mu_{k,s} = M[\overset{\circ}{X}{}^k \overset{\circ}{Y}{}^s]$. Безпосередній розрахунок моментів системи перервних випадкових величин здійснюють за формулами

$$\alpha_{k,s} = \sum_i \sum_j x_i^k y_j^s p_{ij}\,; \qquad \mu_{k,s} = \sum_i \sum_j (x_i - m_x)^k (y_j - m_y)^s p_{ij}\,,$$

де $p_{ij} = p((X = x_i)(Y = y_j))$ – це ймовірність того, що система (X,Y) набуде значення (x_i, y_j), а до суми потрібно долучати усі можливі пари значень величин X та Y.

Крім чисел k та s, які характеризують порядок моменту по відношенню до окремих величин системи, використовують також змішаний (сумарний) порядок моменту – він дорівнює сумі $k+s$ показників степені при X та Y.

У розв'язанні практичних задач використовують тільки моменти першого та другого порядків.

Початкові моменти першого порядку – це **математичні сподівання випадкових величин X та Y**:

$$\alpha_{1,0} = M[X^1 Y^0] = M[X] = m_x\,; \qquad \alpha_{0,1} = M[X^0 Y^1] = M[Y] = m_y\,.$$

Сукупність значень m_x та m_y характеризує положення центра розподілу системи (X,Y).

Центральні моменти другого порядку – це **дисперсії випадкових величин X та Y**:

$$\mu_{2,0} = M[\overset{\circ}{X}{}^2 \overset{\circ}{Y}{}^0] = M[\overset{\circ}{X}{}^2] = D_x\,; \qquad \mu_{0,2} = M[\overset{\circ}{X}{}^0 \overset{\circ}{Y}{}^2] = M[\overset{\circ}{Y}{}^2] = D_y\,.$$

D_x та D_y осібно характеризують розсіювання величин X та Y відносно центрів їх розподілу m_x чи m_y.

Особливе практичне значення має другий змішаний центральний момент

$$\mu_{1,1} = M[\overset{\circ}{X}\,\overset{\circ}{Y}] = K_{xy}\,.$$

K_{xy} називається **кореляційний момент випадкових величин X та Y**:

$$K_{xy} = \sum_i \sum_j (x_i - m_x)(y_j - m_y) p_{ij} .$$

Кореляційний момент K_{xy} – це числова характеристика системи, яка крім розсіювання величин X та Y описує також зв'язок між ними.

Якщо X та Y – незалежні випадкові величини, то кореляційний момент дорівнює нулю. Проте якщо K_{xy} двох величин не дорівнює нулю, то це може бути лише ознакою їх залежності. Адже якщо одна з величин системи практично не відрізняється від свого математичного сподівання, тобто майже не випадкова, то якою б тісною залежністю не були зв'язані між собою величини X та Y, кореляційний момент K_{xy} буде практично дорівнювати нулю. З цієї причини для оцінювання зв'язку між величинами X та Y використовують числову характеристику, яка враховує таку обставину і виражається як нормоване значення кореляційного моменту. Це безрозмірна числова характеристика системи, яка називається **коефіцієнт кореляції** r_{xy}:

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y} .$$

$\sigma_x = \sqrt{D_x}$, $\sigma_y = \sqrt{D_y}$ – середні квадратичні відхилення величин X та Y. Коефіцієнт кореляції незалежних величин завжди дорівнює нулю.

Коефіцієнт кореляції характеризує лінійну імовірнісну залежність випадкових величин. Сутність лінійної імовірнісної залежності полягає в тому, що при збільшенні чи зменшенні однієї величини друга має тенденцію збільшуватись чи зменшуватись з лінійною закономірністю. Така тенденція гранично наближається до строгої лінійної функціональної залежності. Таким чином, r_{xy} характеризує ступінь тісноти лінійної імовірнісної залежності. Якщо величини зв'язані функціональною лінійною залежністю, то $r_{xy} = \pm 1$. Якщо величини зв'язані довільною імовірнісною залежністю, то $-1 < r_{xy} < 1$. За умови, що $r_{xy} > 0$, кореляція величин X та Y додатна, тобто при збільшенні однієї величини друга має тенденцію збільшуватись; якщо $r_{xy} < 0$ – кореляція від'ємна (при збільшенні однієї величини друга має тенденцію зменшуватись). Чим більшого за абсолютною величиною значення набуває коефіцієнт кореляції, тим тісніший зв'язок, тим більше величини залежні між собою. Чим ближче значення коефіцієнта наближається до ± 1, тим більше імовірнісна залежність наближається до функціональної.

Якщо задано результати випробувань $(x_1, y_1);(x_2, y_2);...;(x_n, y_n)$ над величинами системи (X, Y), то про наявність кореляції між ними можна сформулювати попередній висновок на основі зовнішнього вигляду точкової діаграми, де точками відображено одержані за результатами випробувань

пари значень величин X та Y. Типові приклади точкових діаграм для різних видів імовірнісної залежності ілюструють схеми на рис. 8. За умови розташування точок на діаграмі з такою закономірністю, яка дає підстави попередньо вважати зв'язок існуючим, тіснота цього зв'язку може бути виражена статистичним значенням коефіцієнта кореляції.

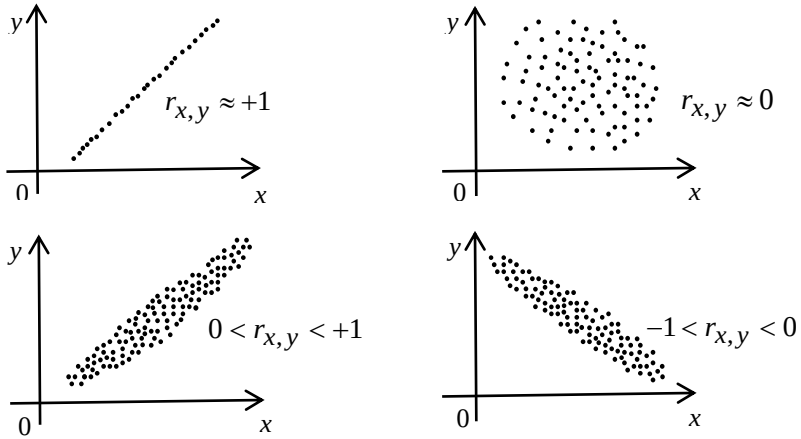


Рис. 8. Приклади точкових діаграм для різних видів імовірнісної залежності

Визначення статистичного значення коефіцієнта кореляції передбачає, у тому числі, розрахунок інших статистичних числових характеристик. **Статистичні числові характеристики** зберігають зміст своїх теоретичних аналогів, а робочі розрахункові формули мають наступний вигляд. Формули розрахунку статистичних математичних сподівань m_x^* та m_y^*, дисперсій D_x^* та D_y^*, середніх квадратичних відхилень σ_x^* та σ_y^*, кореляційного моменту K_{xy}^* і коефіцієнта кореляції r_{xy}^*:

$$m_x^* = \frac{\sum_{i=1}^{n} x_i}{n}; \quad m_y^* = \frac{\sum_{i=1}^{n} y_i}{n}; \quad D_x^* = \frac{\sum_{i=1}^{n} (x_i - \tilde{m}_x)^2}{n}; \quad D_y^* = \frac{\sum_{i=1}^{n} (y_i - \tilde{m}_y)^2}{n};$$

$$\sigma_x^* = \sqrt{D_x^*}; \quad \sigma_y^* = \sqrt{D_y^*}; \quad K_{xy}^* = \frac{\sum_{i=1}^{n} (x_i - \tilde{m}_x)(y_i - \tilde{m}_y)}{n}; \quad r_{xy}^* = \frac{K_{xy}^*}{\sigma_x^* \sigma_y^*}.$$

За умови великої кількості n результатів випробувань статистичні числові характеристики з достатньою точністю виражають відповідні їм ознаки

випадкових величин системи. В умовах обмеженої кількості випробувань статистичні характеристики втрачають здатність точно передавати відповідні ознаки. В такому випадку їх замінюють наближеними значеннями, які називаються **числові оцінки характеристик**. З метою розрахунку числових оцінок математичних сподівань $\tilde{m}_x, \tilde{m}_y$, дисперсій $\tilde{D}_x, \tilde{D}_y$, середніх квадратичних відхилень $\tilde{\sigma}_x, \tilde{\sigma}_y$, кореляційного моменту $\tilde{K}_{xy}$ і коефіцієнта кореляції $\tilde{r}_{xy}$, які задовольняють вимоги найбільшої надійності, незміщеності та ефективності, використовують наступні робочі формули:

$$\tilde{m}_x = m_x^*; \qquad \tilde{m}_y = m_y^*; \qquad \tilde{D}_x = \frac{n}{n-1} D_x^*; \quad \tilde{D}_y = \frac{n}{n-1} D_y^*;$$

$$\tilde{\sigma}_x = \sqrt{\tilde{D}_x}; \qquad \tilde{\sigma}_y = \sqrt{\tilde{D}_y}; \qquad \tilde{K}_{xy} = \frac{n}{n-1} K_{xy}^*; \quad \tilde{r}_{xy} = \frac{\tilde{K}_{xy}}{\tilde{\sigma}_x \tilde{\sigma}_y}.$$

Розрахунок статистичних числових характеристик або їх заміна відповідними числовими оцінками спрямований на визначення тісноти імовірнісного зв’язку. Тісноту імовірнісного зв’язку виражає емпіричне значення коефіцієнта кореляції. Розрахунок коефіцієнта кореляції супроводжується перевіркою його надійності. Алгоритм перевірки неоднозначний і залежить від кількості *n* результатів випробувань.

За умови, що $n \geq 50$, вираженої значенням статистичного коефіцієнта кореляції r_{xy}^* тісноти достатньо для підтвердження наявності імовірнісного зв’язку, якщо одночасно задовольняються вимоги нерівностей

$$\left.\begin{aligned} \left|r_{xy}^*\right| &\geq r_{\min} \\ \left|r_{xy}^*\right| &\geq 3\sigma_r \end{aligned}\right\}.$$

$r_{\min} = \dfrac{\sqrt{n+36} - \sqrt{n}}{6}$ – мінімальне допустиме значення і $\sigma_r = \dfrac{1 - r_{xy}^{*\,2}}{\sqrt{n}}$ – середнє квадратичне відхилення коефіцієнта кореляції.

За умови, що $n < 50$, можна лише наближено схарактеризувати тісноту зв’язку, використовуючи числову оцінку коефіцієнта кореляції $\tilde{r}_{xy}$. Імовірнісний зв’язок між величинами *X* та *Y* підтверджується, якщо одночасно задовольняються вимоги нерівностей

$$\left.\begin{aligned} \left|\tilde{r}_{xy}\right| &\geq r_{\min} \\ \left|\tilde{r}_{xy}\right| &\geq r_2 - r_1 \end{aligned}\right\}.$$

Різниця $r_2 - r_1$ – це довжина довірчого інтервалу $I_\beta = (r_1; r_2)$ для оцінки $\tilde{r}_{xy}$ при довірчій ймовірності $\beta \to 1$. Довірчий інтервал $I_\beta = (r_1; r_2)$ визначається за посередництва функції Фішера, яка підпорядковується нормальному законові розподілу, вигляду

$$Z = \frac{1}{2}(\ln(1+|\tilde{r}_{xy}|) - \ln(1-|\tilde{r}_{xy}|)).$$

Для значень функції Z складено таблицю (див. додаток 6). За значенням оцінки $\tilde{r}_{xy}$ виражається оцінка $\tilde{Z}$ і визначається її довірчий інтервал $I_\beta^{(Z)} = (Z_1; Z_2)$, де $Z_1 = \tilde{Z} - t_\beta \sigma_{\tilde{Z}}$; $Z_2 = \tilde{Z} + t_\beta \sigma_{\tilde{Z}}$; $t_\beta = \arg \Phi^*(\frac{1+\beta}{2})$ – аргумент нормальної функції розподілу для заданої довірчої ймовірності β; $\sigma_{\tilde{Z}} = \frac{1}{\sqrt{n-3}}$ – середнє квадратичне відхилення оцінки $\tilde{Z}$. Границі r_1 та r_2 довірчого інтервалу $I_\beta = (r_1; r_2)$ визначаються за відповідними границями Z_1 та Z_2 довірчого інтервалу $I_\beta^{(Z)} = (Z_1; Z_2)$ із розв'язування оберненої задачі на основі функції Z чи таблиці її значень.

За умови підтвердженої імовірнісної залежності в системі (X, Y), її виражають аналітичною формою у вигляді лінійної емпіричної формули. Емпірична формула, яка виражає лінійний імовірнісний зв'язок в системі (X, Y), називається **рівняння регресії**. Рівняння регресії мають вигляд

$$y_i - m_y = \rho_{y/x}(x_i - m_x), \qquad x_i - m_x = \rho_{x/y}(y_i - m_y)$$

або

$$y_i = \rho_{y/x} x_i + (m_y - \rho_{y/x} m_x) \qquad x_i = \rho_{x/y} y_i + (m_x - \rho_{x/y} m_y).$$

Тут $\rho_{y/x} = r_{xy}\frac{\sigma_y}{\sigma_x}$ – коефіцієнт регресії Y на X; $\rho_{x/y} = r_{xy}\frac{\sigma_x}{\sigma_y}$ – коефіцієнт регресії X на Y.

Рівняння регресії – це рівняння прямої, яка апроксимує усі результати випробувань. Рівняння регресії дають змогу розраховувати ймовірні значення величини Y, якщо задатись значенням X, і навпаки, виходячи з тісноти підтвердженої лінійної імовірнісної залежності між цими величинами.

Хід роботи

Під час виконання роботи проводиться кореляційний аналіз результатів геодезичних вимірів, які складають систему двох випадкових величин.

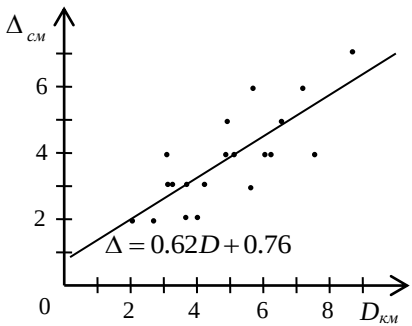
Завдання. В результаті n=20 проведених польових вимірів довжин сторін мережі трилатерації одержано значення D виміряних довжин і відповідні кожній довжині похибки Δ. Необхідно визначити тісноту і форму зв'язку в системі випадкових величин (D, Δ).

Вхідні дані для виконання завдання зведено до таблиці додатку 8. З таблиці необхідно вибрати підряд n=20 значень величин D і Δ, розпочинаючи з номера варіанта N.

Приклад розв'язування завдання. Допустимо, задано наступні пари значень величин X=D і Y=Δ:

№	$D_{(км)}$	$\Delta_{(см)}$	№	$D_{(км)}$	$\Delta_{(см)}$	№	$D_{(км)}$	$\Delta_{(см)}$	№	$D_{(км)}$	$\Delta_{(см)}$
1	8.7	7.0	6	6.1	4.0	11	5.7	6.0	16	2.0	2.0
2	3.7	3.0	7	2.7	3.0	12	4.9	5.0	17	4.0	2.0
3	6.0	4.0	8	4.9	4.0	13	5.6	3.0	18	6.5	5.0
4	3.3	3.0	9	3.1	4.0	14	7.6	4.0	19	7.2	6.0
5	5.1	4.0	10	3.7	2.0	15	4.2	3.0	20	2.7	2.0

1. На початковій стадії вирішення завдання будуємо точкову діаграму залежності довжин D і похибок Δ. Вздовж осі абсцис відкладаємо значення довжин, вздовж осі ординат – відповідні їм значення похибок, а на перетині ставимо точки.



Закономірність взаємного розташування точок на діаграмі вказує на виразну додатну кореляцію між величинами D і Δ. Це є підставою для виконання наступних дій, які передбачає кореляційний аналіз.

2. Визначаємо тісноту залежності довжин D і похибок Δ. При кількості вимірів $n < 50$ тісноту залежності наближено можна виразити формулою числової оцінки коефіцієнта кореляції $\tilde{r}_{D,\Delta}$, використовуючи для проміжних розрахунків формули числових оцінок математичних сподівань $\tilde{m}_D, \tilde{m}_\Delta$, дисперсій $\tilde{D}_D, \tilde{D}_\Delta$, середніх квадратичних відхилень $\tilde{\sigma}_D, \tilde{\sigma}_\Delta$ і кореляційного моменту $\tilde{K}_{D\Delta}$. Результати проміжних розрахунків розмістимо в таблиці:

Результати вимірів			Результати розрахунків				
№	D_i	Δ_i	$D_i - \tilde{m}_D$	$\Delta_i - \tilde{m}_\Delta$	$(D_i - \tilde{m}_D)^2$	$(\Delta_i - \tilde{m}_\Delta)^2$	$(D_i - \tilde{m}_D)\times(\Delta_i - \tilde{m}_\Delta)$
1	8.7	7.0	+3.8	+3.2	14.44	10.24	12.16
2	3.7	3.0	-1.2	-0.8	1.44	0.64	0.96
3	6.0	4.0	+1.1	+0.2	1.21	0.04	0.22
4	3.3	3.0	-1.6	-0.8	2.56	0.64	1.28
5	5.1	4.0	+0.2	+0.2	0.04	0.04	0.04
6	6.1	4.0	+1.2	+0.2	1.44	0.04	0.24
7	2.7	3.0	-2.2	-0.8	4.84	0.64	1.76
8	4.9	4.0	0.0	+0.2	0.0	0.04	0.00
9	3.1	4.0	-1.8	+0.2	3.24	0.04	-0.36
10	3.7	2.0	-1.2	-1.8	1.44	3.24	2.16
11	5.7	6.0	+0.8	+2.2	0.64	4.84	1.76
12	4.9	5.0	0.0	+1.2	0.0	1.44	0.00
13	5.6	3.0	+0.7	-0.8	0.49	0.64	-0.56
14	7.6	4.0	+2.7	+0.2	7.29	0.04	0.54
15	4.2	3.0	-0.7	-0.8	0.49	0.64	0.56
16	2.0	2.0	-2.9	-1.8	8.41	3.24	5.22
17	4.0	2.0	-0.9	-1.8	0.81	3.24	1.62
18	6.5	5.0	+1.6	+1.2	2.56	1.44	1.92
19	7.2	6.0	+2.3	+2.2	5.29	4.84	5.06
20	2.7	2.0	-2.2	-1.8	4.84	3.24	3.96
Σ	97.7	76.0	-	-	61.47	39.20	38.54

Підстановка результатів проміжних розрахунків у відповідні формули дає наступні результати: $\tilde{m}_D = 4.9$, $\tilde{m}_\Delta = 3.8$, $\tilde{\sigma}_D = 1.80$, $\tilde{\sigma}_\Delta = 1.44$, $\tilde{K}_{D,\Delta} = 2.03$; наближене значення тісноти кореляційного зв'язку $\tilde{r}_{D,\Delta} = 0.78$.

3. Перевіряємо вимоги наявності імовірнісного зв'язку $\left.\begin{array}{l}\left|\tilde{r}_{D,\Delta}\right| \geq r_{\min} \\ \left|\tilde{r}_{D,\Delta}\right| \geq r_2 - r_1\end{array}\right\}$.

Мінімальне допустиме значення числової оцінки коефіцієнта кореляції $r_{\min} = \dfrac{\sqrt{n+36} - \sqrt{n}}{6} = 0.50$. Отже, $\left|\tilde{r}_{D,\Delta}\right| > r_{\min}$ – вимога першої нерівності задовольняється.

Перевірку вимоги другої нерівності здійснюємо шляхом оцінювання точності коефіцієнта $\tilde{r}_{D,\Delta}$ довірчим інтервалом $I_\beta = (r_1; r_2)$ при довірчій ймовірності β = 0.9. За таблицею додатку 6 при $\tilde{r}_{D,\Delta}$ = 0.78 визначаємо значення функції $\tilde{Z}$ = 1.0454. Середнє квадратичне відхилення величини $\tilde{Z}$ $\sigma_{\tilde{Z}} = \frac{1}{\sqrt{n-3}}$ = 0.243. Значення параметра $t_\beta = \arg\Phi^*(\frac{1+\beta}{2})$ визначаємо з таблиці нормальної функції розподілу (додаток 4): $t_\beta = \arg\Phi^*(0.95)$ = 1.645. Величина Z може набувати значень $\tilde{Z} - t_\beta\sigma_{\tilde{Z}} \le Z \le \tilde{Z} + t_\beta\sigma_{\tilde{Z}}$. При довірчій ймовірності 0.9 її довірчий інтервал $I_\beta^{(Z)}$ = (0.6457; 1.4451). За значеннями Z_1 = 0.6457 та Z_2 = 1.4451 шляхом оберненого табличного інтерполювання у таблиці додатку 6 визначаємо границі r_1 =0.57 та r_2 =0.89 довірчого інтервалу для коефіцієнта кореляції $I_\beta = (r_1; r_2)$. З надійністю β = 0.9 точність невідомого значення коефіцієнта кореляції $r_{D,\Delta}$ окреслюється границями довірчого інтервалу I_β = (0.57; 0.89). Довжина інтервалу $r_2 - r_1$ = 0.32. Отже, $\left|\tilde{r}_{D,\Delta}\right| > r_2 - r_1$ – вимога другої нерівності також задовольняється.

Таким чином, наявність лінійного імовірнісного зв'язку між довжинами D та похибками Δ підтверджена.

4. Підтвердження імовірнісного зв'язку між довжинами та їх похибками є підставою виразити його аналітичною формою у вигляді лінійного рівняння регресії. В розв'язуваному завданні є зміст складати лише рівняння регресії Δ на D вигляду $\Delta_i = \rho_{\Delta/D} D_i + (\tilde{m}_\Delta - \rho_{\Delta/D}\tilde{m}_D)$. Розраховуємо коефіцієнт регресії $\rho_{\Delta/D} = \tilde{r}_{D,\Delta}\frac{\tilde{\sigma}_\Delta}{\tilde{\sigma}_D}$ = 0.62 і вільний член рівняння $\tilde{m}_\Delta - \rho_{\Delta/D}\tilde{m}_D$ = 0.76. Отже, рівняння регресії має вигляд $\Delta_i = 0.62 D_i + 0.76$. Рівняння виражає лінійну імовірнісну залежність похибок Δ від довжин D, одержаних за відповідного комплексу умов. Воно дає можливість розраховувати ймовірні похибки вимірів ліній для довільних значень довжин за умови збереження того ж комплексу умов. Рівняння регресії D на Δ формально може бути побудоване, однак задача розрахунку довжини лінії за похибкою її виміру позбавлене логічного змісту.

Рівняння регресії відображуємо графічно у вигляді прямої на одному рисунку з точковою діаграмою залежності довжин D і похибок Δ.

Питання для самоконтролю

1. Який зміст поняття лінійної імовірнісної залежності в системі двох випадкових величин?

2. Якими числовими характеристиками описується система випадкових величин? Який зміст мають ці характеристики?
3. Що характеризує і як обчислюється статистичний коефіцієнт кореляції?
4. Коли і як обчислюється числова оцінка коефіцієнта кореляції?
5. Як визначити точність та надійність числової оцінки коефіцієнта кореляції?
6. Як можна зробити попередній висновок про наявність імовірнісного зв'язку між випадковими величинами?
7. Як виразити тісноту лінійної імовірнісної залежності в системі двох випадкових величин?
8. Як підтверджується наявність імовірнісного зв'язку між випадковими величинами при кількості випробувань $n \geq 50$?
9. Як підтверджується наявність імовірнісного зв'язку між випадковими величинами при кількості випробувань $n < 50$?
10. Що таке рівняння регресії? Як записати таке рівняння?
11. Що виражає рівняння регресії і як воно використовується на практиці?
12. Як пояснити зміст термінів „тіснота” та „форма” імовірнісного зв'язку в системі двох випадкових величин?

Література

1. Вентцель Е. С. Теория вероятностей : підручник. Москва : Наука, 1969. 576 с.
2. Видуев Н. Г., Григоренко А. Г. Математическая обработка геодезических измерений : навчальний посібник. Київ : Вища школа, 1978. 376 с.
3. Войтенко С. П. Математична обробка геодезичних вимірів. Теорія похибок вимірів : навчальний посібник. Київ : КНУБА, 2003. 216 с.
4. Основи математичного опрацювання геодезичних вимірів : підручник / Зазуляк П. М., Гавриш В. І., Євсєєва Е. М., Йосипчук М. Д. Львів : Растр-7, 2007. 408 с.

Додаток 1

Значення χ^2 залежно від r і p

r \ p	0.99	0.98	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.02	0.01	0.001
1	0.000	0.001	0.004	0.016	0.064	0.148	0.455	1.074	1.642	2.71	3.84	5.41	6.64	10.83
2	0.020	0.040	0.103	0.211	0.446	0.713	1.386	2.41	3.22	4.60	5.99	7.82	9.21	13.82
3	0.115	0.185	0.352	0.584	1.005	1.424	2.37	3.66	4.64	6.25	7.82	9.84	11.34	16.27
4	0.297	0.429	0.711	1.064	1.649	2.20	3.36	4.88	5.99	7.78	9.49	11.67	13.28	18.46
5	0.554	0.752	1.145	1.610	2.34	3.00	4.35	6.06	7.29	9.24	11.07	13.39	15.09	20.5
6	0.872	1.134	1.635	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	15.03	16.81	22.5
7	1.239	1.564	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	16.62	18.48	24.3
8	1.646	2.03	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	18.17	20.1	26.1
9	2.09	2.53	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	19.68	21.7	27.9
10	2.56	3.06	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	21.2	23.2	29.6
11	3.05	3.61	4.58	5.58	6.99	8.15	10.34	12.90	14.63	17.28	19.68	22.6	24.7	31.3
12	3.57	4.18	5.23	6.30	7.81	9.03	11.34	14.01	15.81	18.55	21.0	24.1	26.2	32.9
13	4.11	4.76	5.89	7.04	8.63	9.93	12.34	15.12	16.98	19.81	22.4	25.5	27.7	34.6
14	4.66	5.37	6.57	7.79	9.47	10.82	13.34	16.22	18.15	21.1	23.7	26.9	29.1	36.1
15	5.23	5.98	7.26	8.55	10.31	11.72	14.34	17.32	19.31	22.3	25.0	28.3	30.6	37.7
16	5.81	6.61	7.96	9.31	11.15	12.62	15.34	18.42	20.5	23.5	26.3	29.6	32.0	39.3
17	6.41	7.26	8.67	10.08	12.00	13.53	16.34	19.51	21.6	24.8	27.6	31.0	33.4	40.8
18	7.02	7.91	9.39	10.86	12.86	14.44	17.34	20.6	22.8	26.0	28.9	32.3	34.8	42.3
19	7.63	8.57	10.11	11.65	13.72	15.35	18.34	21.7	23.9	27.2	30.1	33.7	36.2	43.8
20	8.26	9.24	10.85	12.44	14.58	16.27	19.34	22.8	25.0	28.4	31.4	35.0	37.6	45.3
21	8.90	9.92	11.59	13.24	15.44	17.18	20.3	23.9	26.2	29.6	32.7	36.3	38.9	46.8
22	9.54	10.60	12.34	14.04	16.31	18.10	21.3	24.9	27.3	30.8	33.9	37.7	40.3	48.3
23	10.20	11.29	13.09	14.85	17.19	19.02	22.3	26.0	28.4	32.0	35.2	39.0	41.6	49.7
24	10.86	11.99	13.85	15.66	18.06	19.94	23.3	27.1	29.6	33.2	36.4	40.3	43.0	51.2
25	11.52	12.70	14.61	16.47	18.94	20.9	24.3	28.2	30.7	34.4	37.7	41.7	44.3	52.6
26	12.20	13.41	15.38	17.29	19.82	21.8	25.3	29.2	31.8	35.6	38.9	42.9	45.6	54.1
27	12.88	14.12	16.15	18.11	20.7	22.7	26.3	30.3	32.9	36.7	40.1	44.1	47.0	55.5
28	13.56	14.85	16.93	18.94	21.6	23.6	27.3	31.4	34.0	37.9	41.3	45.4	48.3	56.9
29	14.26	15.57	17.71	19.77	22.5	24.6	28.3	32.5	35.1	39.1	42.6	46.7	49.6	58.3
30	14.95	16.31	18.49	20.6	23.4	25.5	29.3	33.5	36.2	40.3	43.8	48.0	50.9	59.7

Додаток 2

Значення ймовірностей $p(\lambda)$

λ	$p(\lambda)$	λ	$p(\lambda)$	λ	$p(\lambda)$
0.0	1.000	0.7	0.711	1.4	0.040
0.1	1.000	0.8	0.544	1.5	0.022
0.2	1.000	0.9	0.393	1.6	0.012
0.3	1.000	1.0	0.270	1.7	0.006
0.4	0.997	1.1	0.178	1.8	0.003
0.5	0.964	1.2	0.112	1.9	0.002
0.6	0.864	1.3	0.068	2.0	0.001

Додаток 3

Значення функції $f(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$.

t	0	1	2	3	4	5	6	7	8	9	t
0.0	0.3989	0.3989	0.3989	0.3988	0.3986	0.3984	0.3982	0.3980	0.3977	0.3973	0.0
0.1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918	0.1
0.2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825	0.2
0.3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697	0.3
0.4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538	0.4
0.5	0.3521	0.3503	0.3485	0.3467	0.3448	0.3429	0.3410	0.3391	0.3372	0.3352	0.5
0.6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144	0.6
0.7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920	0.7
0.8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685	0.8
0.9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444	0.9
1.0	0.2420	0.2396	0.2371	0.2347	0.2323	0.2299	0.2275	0.2251	0.2227	0.2203	1.0
1.1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965	1.1
1.2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736	1.2
1.3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518	1.3
1.4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315	1.4
1.5	0.1295	0.1276	0.1257	0.1238	0.1219	0.1200	0.1182	0.1163	0.1145	0.1127	1.5
1.6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957	1.6
1.7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804	1.7
1.8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669	1.8
1.9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551	1.9
2.0	0.0540	0.0529	0.0519	0.0508	0.0498	0.0488	0.0478	0.0468	0.0459	0.0449	2.0
2.1	0440	0431	0422	0413	0404	0396	0388	0379	0371	0363	2.1
2.2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290	2.2
2.3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229	2.3
2.4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180	2.4
2.5	0.0175	0.0171	0.0167	0.0163	0.0158	0.0154	0.0151	0.0147	0.0143	0.0139	2.5
2.6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107	2.6
2.7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081	2.7
2.8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061	2.8
2.9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046	2.9
3.0	0.0044	0.0043	0.0042	0.0040	0.0039	0.0038	0.0037	0.0036	0.0035	0.0034	3.0
3.1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025	3.1
3.2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018	3.2
3.3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013	3.3
3.4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009	3.4
3.5	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007	0.0007	0.0007	0.0006	3.5
3.6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004	3.6
3.7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003	3.7
3.8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002	3.8
3.9	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001	3.9
t	0	1	2	3	4	5	6	7	8	9	t

Додаток 4

Значення нормальної функції розподілу $\Phi^*(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{t^2}{2}} dt$

t	$\Phi^*(t)$	t	$\Phi^*(t)$	t	$\Phi^*(t)$	t	$\Phi^*(t)$	t	$\Phi^*(t)$	t	$\Phi^*(t)$
0.00	0.5000	0.40	0.6554	0.80	0.7881	1.20	0.8849	1.60	0.9452	2.00	0.9772
0.01	5040	0.41	6591	0.81	7910	1.21	8869	1.61	9463	2.10	9821
0.02	5080	0.42	6628	0.82	7939	1.22	8888	1.62	9474	2.20	9861
0.03	5120	0.43	6664	0.83	7967	1.23	8907	1.63	9484	2.30	9893
0.04	5160	0.44	6700	0.84	7995	1.24	8925	1.64	9495	2.40	9918
0.05	5199	0.45	6736	0.85	8023	1.25	8944	1.65	9505	2.50	9938
0.06	5239	0.46	6772	0.86	8051	1.26	8962	1.66	9515	2.60	9953
0.07	5279	0.47	6808	0.87	8078	1.27	8980	1.67	9525	2.70	9965
0.08	5319	0.48	6844	0.88	8106	1.28	8997	1.68	9535	2.80	9974
0.09	5359	0.49	6879	0.89	8133	1.29	9015	1.69	9545	2.90	9981
0.10	0.5398	0.50	0.6915	0.90	0.8159	1.30	0.9032	1.70	0.9554	3.00	0.9986
0.11	5438	0.51	6950	0.91	8186	1.31	9049	1.71	9564	3.10	9990
0.12	5478	0.52	6985	0.92	8212	1.32	9066	1.72	9573	3.20	9993
0.13	5517	0.53	7019	0.93	8238	1.33	9082	1.73	9582	3.30	9995
0.14	5557	0.54	7054	0.94	8264	1.34	9099	1.74	9591	3.40	9997
0.15	5596	0.55	7088	0.95	8289	1.35	9115	1.75	9599	3.50	9998
0.16	5636	0.56	7123	0.96	8315	1.36	9131	1.76	9608	3.60	9998
0.17	5675	0.57	7157	0.97	8340	1.37	9147	1.77	9616	3.70	9999
0.18	5714	0.58	7190	0.98	8365	1.38	9162	1.78	9625	3.80	0.9999
0.19	5753	0.59	7224	0.99	8389	1.39	9177	1.79	9633	3.90	1.0000
0.20	0.5793	0.60	0.7257	1.00	0.8413	1.40	0.9192	1.80	0.9641		
0.21	5832	0.61	7291	1.01	8437	1.41	9207	1.81	9649		
0.22	5871	0.62	7324	1.02	8461	1.42	9222	1.82	9656		
0.23	5910	0.63	7357	1.03	8485	1.43	9236	1.83	9664		
0.24	5948	0.64	7389	1.04	8508	1.44	9251	1.84	9671		
0.25	5987	0.65	7422	1.05	8531	1.45	9265	1.85	9678		
0.26	6026	0.66	7454	1.06	8554	1.46	9279	1.86	9686		
0.27	6064	0.67	7486	1.07	8577	1.47	9292	1.87	9693		
0.28	6103	0.68	7517	1.08	8599	1.48	9306	1.88	9699		
0.29	6141	0.69	7549	1.09	8621	1.49	9319	1.89	9706		
0.30	0.6179	0.70	0.7580	1.10	0.8643	1.50	0.9332	1.90	0.9713		
0.31	6217	0.71	7611	1.11	8665	1.51	9345	1.91	9719		
0.32	6255	0.72	7642	1.12	8686	1.52	9357	1.92	9726		
0.33	6293	0.73	7673	1.13	8708	1.53	9370	1.93	9732		
0.34	6331	0.74	7703	1.14	8729	1.54	9382	1.94	9738		
0.35	6368	0.75	7734	1.15	8749	1.55	9394	1.95	9744		
0.36	6406	0.76	7764	1.16	8770	1.56	9406	1.96	9750		
0.37	6443	0.77	7794	1.17	8790	1.57	9418	1.97	9756		
0.38	6480	0.78	7823	1.18	8810	1.58	9429	1.98	9761		
0.39	0.6517	0.79	0.7852	1.19	0.8830	1.59	0.9441	1.99	0.9767		

Додаток 5

Значення параметра розподілу Стьюдента t_β залежно від β та n-1

n-1 \ β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.98	0.99	0.999
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.08	6.31	12.71	31.8	63.7	636.6
2	142	289	445	617	0.816	1.061	336	1.886	2.92	4.30	6.96	9.92	31.6
3	137	277	424	584	765	0.978	250	638	2.35	3.18	4.54	5.84	12.94
4	134	271	414	569	741	941	190	533	2.13	2.77	3.75	4.60	8.61
5	132	267	408	559	727	920	156	476	2.02	57	3.36	4.03	6.86
6	132	265	404	553	718	906	134	440	1.943	45	3.14	3.71	5.96
7	130	263	402	549	711	896	119	415	895	36	3.00	50	5.40
8	130	262	399	546	706	889	108	397	860	31	2.90	36	5.04
9	129	261	398	543	703	883	100	383	833	26	82	25	4.78
10	129	260	397	542	700	879	093	372	812	23	76	17	4.59
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.20	2.72	3.11	4.49
12	128	259	395	539	695	873	083	356	782	18	68	3.06	32
13	128	259	394	538	694	870	079	350	771	16	65	3.01	22
14	128	258	393	537	692	868	076	345	761	14	62	2.98	14
15	128	258	393	536	691	866	074	341	753	13	60	95	07
16	128	258	392	535	690	865	071	337	746	12	58	92	4.02
17	128	257	392	534	689	863	069	333	740	11	57	90	3.96
18	127	257	392	534	688	862	067	330	734	10	55	88	92
19	127	257	391	533	688	861	066	328	729	09	54	86	88
20	127	257	391	533	687	860	064	325	725	09	53	84	85
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.08	2.52	2.83	3.82
22	127	256	390	532	686	858	061	321	717	07	51	82	79
23	127	256	390	532	685	858	060	319	714	07	50	81	77
24	127	256	390	531	685	857	059	318	711	06	49	80	74
25	127	256	390	531	684	856	058	316	708	06	48	79	72
26	127	256	390	531	684	856	058	315	706	06	48	78	71
27	127	256	389	531	684	855	057	314	703	05	47	77	69
28	127	256	389	530	683	855	056	313	701	05	47	76	67
29	127	256	389	530	683	854	055	311	699	04	46	76	66
30	127	256	389	530	683	854	055	310	697	04	46	75	65
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.02	2.42	2.70	3.55
60	126	254	387	527	679	848	046	296	671	2.00	39	66	46
120	126	254	386	526	677	845	041	289	658	1.98	36	62	37
∞	0.126	0.253	0.385	0.524	0.674	0.842	1.036	1.282	1.645	1.96	2.33	2.58	3.29

Додаток 6

Значення функції $Z = \frac{1}{2} \ln \frac{1+r}{1-r}$

r	0	1	2	3	4	5	6	7	8	9	*r*
0.0	0.0000	0.0100	0.0200	0.0300	0.0400	0.0500	0.0601	0.0701	0.0802	0.0902	0.0
0.1	0.1003	0.1104	0.1206	0.1307	0.1409	0.1511	0.1614	0.1717	0.1820	0.1923	0.1
0.2	0.2027	0.2132	0.2237	0.2342	0.2448	0.2554	0.2661	0.2769	0.2877	0.2986	0.2
0.3	0.3095	0.3205	0.3316	0.3428	0.3541	0.3654	0.3769	0.3881	0.4001	0.4118	0.3
0.4	0.4236	0.4356	0.4477	0.4599	0.4722	0.4847	0.4973	0.5101	0.5230	0.5361	0.4
0.5	0.5493	0.5627	0.5763	0.5901	0.6042	0.6184	0.6328	0.6475	0.6625	0.6777	0.5
0.6	0.6931	0.7089	0.7250	0.7414	0.7582	0.7753	0.7928	0.8107	0.8291	0.8480	0.6
0.7	0.8673	0.8872	0.9076	0.9287	0.9505	0.9730	0.9962	1.0203	1.0454	1.0714	0.7
0.8	1.0986	1.1270	1.1568	1.1881	1.2212	1.2562	1.2933	1.3331	1.3758	1.4219	0.8
0.9	1.4722	1.5275	1.5890	1.6584	1.7380	1.8318	1.9459	2.0923	2.2976	2.6466	0.9
0.99	2.6466	2.6996	2.7587	2.8257	2.9031	2.9945	3.1063	3.2504	3.4534	3.8002	0.99

Додаток 7

Вхідні дані для виконання завдання теми 1

(m_i – кількість похибок в інтервалах статистичного ряду розподілу)

№ варіанта	-4;-3	-3;-2	-2;-1	-1;0	0; 1	1; 2	2; 3	3; 4	№ варіанта	-4;-3	-3;-2	-2;-1	-1;0	0; 1	1; 2	2; 3	3; 4
1	13	48	91	152	135	76	45	11	26	9	24	44	68	62	50	32	12
2	10	49	78	145	130	75	49	12	27	8	20	31	48	59	45	25	9
3	9	44	73	120	138	88	33	9	28	10	15	30	57	50	39	21	17
4	10	29	71	140	128	79	34	12	29	8	20	34	54	62	48	19	10
5	11	44	89	125	143	82	34	8	30	8	18	29	54	51	40	31	12
6	14	51	88	151	140	73	40	11	31	10	24	54	83	89	38	27	15
7	15	29	84	134	145	81	35	17	32	16	35	74	129	136	84	38	14
8	10	44	77	119	129	80	27	16	33	14	32	90	138	129	82	42	13
9	11	25	48	88	92	34	26	12	34	17	36	72	138	128	83	40	12
10	9	22	43	76	69	36	24	8	35	23	90	131	145	129	84	39	10
11	10	24	48	85	71	33	25	9	36	12	31	86	125	119	83	40	15
12	10	25	52	91	94	35	24	12	37	15	48	87	150	132	78	51	14
13	9	21	54	96	91	30	20	12	38	10	27	79	124	131	80	32	9
14	9	23	56	90	84	28	19	10	39	11	41	69	130	118	68	25	13
15	17	44	91	130	142	80	38	20	40	16	52	92	161	140	81	49	15
16	16	39	95	164	172	87	43	7	41	12	45	77	120	134	80	33	10
17	12	31	89	150	140	86	41	16	42	10	23	49	88	80	39	25	11
18	9	42	82	121	138	92	38	10	43	9	14	78	135	146	80	35	15
19	8	34	85	132	119	76	30	11	44	8	23	59	98	89	36	28	11
20	10	29	75	115	108	66	40	15	45	16	40	88	129	142	106	78	22
21	9	40	61	96	107	59	31	8	46	14	70	132	182	169	91	59	11
22	11	32	68	101	96	71	42	16	47	13	64	120	166	178	120	54	12
23	19	36	61	86	93	52	28	15	48	12	58	122	160	151	95	69	16
24	12	29	54	84	77	59	30	18	49	11	38	69	135	154	109	50	13
25	9	25	49	69	76	48	21	10	50	10	34	89	142	130	96	58	15

Додаток 8

Вхідні дані для виконання завдання теми 3

№	$D_{(км)}$	$\Delta_{(см)}$	№	$D_{(км)}$	$\Delta_{(см)}$	№	$D_{(км)}$	$\Delta_{(см)}$	№	$D_{(км)}$	$\Delta_{(см)}$
1	8.7	7	16	8.4	7	31	6.3	5	46	3.8	3
2	5.4	3	17	2.7	3	32	4.7	4	47	7.2	6
3	3.7	3	18	3.2	3	33	3.5	3	48	5.6	5
4	6.8	4	19	4.9	4	34	5.6	3	49	5.1	4
5	7.7	5	20	4.6	3	35	7.6	4	50	2.7	2
6	6.0	4	21	3.1	4	36	3.9	4	51	8.1	6
7	3.2	4	22	2.7	2	37	4.2	3	52	5.6	4
8	2.5	3	23	4.1	3	38	6.6	5	53	3.8	3
9	3.3	3	24	3.7	2	39	3.7	3	54	6.8	5
10	6.5	4	25	3.9	3	40	2.0	2	55	7.7	6
11	5.1	4	26	7.4	6	41	4.1	3	56	6.1	5
12	5.8	4	27	5.7	6	42	8.3	7	57	3.7	4
13	7.9	6	28	5.0	4	43	4.0	2	58	2.7	3
14	6.1	4	29	8.9	7	44	7.8	6	59	3.1	3
15	7.2	6	30	4.9	5	45	6.5	5	60	6.8	5