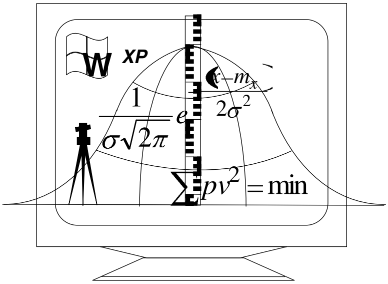
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ВОДНОГО ГОСПОДАРСТВА ТА ПРИРОДОКОРИСТУВАННЯ

Інститут агроекології та землеустрою

Кафедра геодезії та геоінформатики



05-04-31

МЕТОДИЧНІ ВКАЗІВКИ

до виконання самостійних та практичних робіт з дисципліни

„МАТЕМАТИЧНА ОБРОБКА ГЕОДЕЗИЧНИХ ВИМІРІВ”

студентами напряму підготовки 0801 „Геодезія, картографія та землеустрій”

Розділ 1.

Елементи математичної статистики

Рекомендовано
методичною комісією напряму підготовки 0801
„Геодезія, картографія та землеустрій”.
Протокол № 2 від 22 жовтня 2013р.

РІВНЕ – 2014

Методичні вказівки до виконання самостійних та практичних робіт з дисципліни „Математична обробка геодезичних вимірів” студентами напряму підготовки 0801 „Геодезія, картографія та землеустрій” Розділ 1. Елементи математичної статистики / О.А.Тадєєв, Т.І. Дець, Рівне: НУВГП, 2014. – 37 с.

Упорядники: О.А. Тадєєв, кандидат технічних наук, доцент кафедри геодезії та геоінформатики;
Т.І. Дець, асистент кафедри геодезії та геоінформатики.

ЗМІСТ

сторінка

ВСТУП

Всяке дослідження випадкових явищ в теорії ймовірностей прямо чи посередньо опирається на результати спостережень, вимірів чи інші експериментальні дані. Розробкою методів реєстрації, накопичення, описування та аналізу статистичних експериментальних даних для дослідження масових випадкових явищ займається інша математична наука – математична статистика. Основні завдання математичної статистики – це визначення закону розподілу випадкової величини, перевірка правдоподібності теоретичних гіпотез і розрахунок невідомих параметрів розподілу величини. Всі завдання математичної статистики зводяться до обробки результатів спостережень масових випадкових явищ. Залежно від характеру та змісту поставленої практичної задачі та об'єму експериментальних даних ці статистичні завдання набувають різних форм. Різні форми статистичних завдань і є тим математичним апаратом, яким реалізуються в конкретних практичних завданнях теореми й закони теорії ймовірностей.

Розробка методів реєстрації, описання та аналізу дослідних даних, які накопичуються в результаті спостереження масових випадкових явищ, складає предмет математичної статистики.

1. ВИЗНАЧЕННЯ ЗАКОНУ РОЗПОДІЛУ НА ОСНОВІ ДОСЛІДНИХ ДАНИХ

Основні теоретичні положення

Найбільш важливе завдання математичної статистики - визначення закону розподілу випадкової величини на основі певної сукупності значень, яких вона набуває в результаті n проведених незалежних випробувань. Сукупність таких значень величини називається простою статистичною сукупністю і є первинним статистичним матеріалом, який підлягає обробці та аналізу. Просту статистичну сукупність оформлюють у вигляді таблиці, в першому стовпці якої вказують номер випробування, а в другому – його результат. При зростанні числа випробувань проста статистична сукупність з огляду на її громіздкість стає незручною формою запису статистичного матеріалу. Тому її перетворюють до більш зручного компактного вигляду. Для цього весь діапазон спостережених значень випадкової величини ділять на k інтервалів і рахують числа m_i значень величини, які містяться в кожному інтервалі (i=1,2,...,k). Якщо величина набуває значень, які дорівнюють границям вибраних інтервалів, то такі значення слід відносити до інтервалів, у яких ліва границя дорівнює відповідному значенню величини. Контролем підрахунку m_i є їх сума: $\sum_{i=1}^{k} m_i = n$. Значення m_i, поділене на загальне число випробувань n, дає відносну частоту (або статистичну ймовірність) попадання величини в кожний i-ий інтервал:

$$p_i^* = Q = \frac{m_i}{n}. \qquad (1.1.1)$$

Сума всіх отриманих частот повинна дорівнювати одиниці. Таблиця, в якій наведені інтервали в порядку їх розташування вздовж осі абсцис та відповідні кожному інтервалові відносні частоти, називається статистичним рядом розподілу.

З практики встановлено, що число інтервалів k раціонально вибирати порядку 10 і більше при відповідно зростаючому об’ємі статистичного матеріалу, однак за умови, щоб значення m_i були достатньо великими з імовірнісної точки зору (рекомендуються значення, не менші 5÷10). Інтервали можуть бути як однакової, так і різної довжини.

Графічне оформлення статистичного ряду розподілу називається гістограмою. Для побудови гістограми на осі абсцис позначають інтервали і на основі кожного інтервалу будують прямокутник з площею, яка дорівнює частоті інтервалу. Висотою прямокутника є частота інтервалу, поділена на його довжину. Повна площа гістограми дорівнює одиниці.

Одним із способів обробки дослідних даних є побудова статистичної функції розподілу випадкової величини. Статистичною функцією розподілу

$F^*(x)$ випадкової величини X називається частота події $X < x$ в даному статистичному матеріалі:
$$F^*(x) = P^*(X < x). \tag{1.1.2}$$
Для визначення функції $F^*(x)$ при потрібному значенні змінної x рахують число випробувань, в яких величина набула значень, які менші від x, і ділять його на загальне число випробувань. Статистична функція розподілу є перервною ступінчатою функцією. Розриви відповідають спостереженим значенням величини і за величиною дорівнюють частотам цих значень. Отримана таким способом функція буде достатньо точно описувати дослідні дані, однак її побудова при великому обсязі статистичного матеріалу і присвоюванні змінній x послідовно всіх значень величини перетворюється в дуже трудомісткий процес. Тому статистичну функцію розподілу можна побудувати приблизно за даними статистичного ряду розподілу, вибираючи по черзі фіксованими значеннями змінної величини x границі інтервалів. Тоді
$$F^*(x) = \sum_{x_i < x} p_i^*, \tag{1.1.3}$$
де i=1,2,..., k; k - число інтервалів; символ $x_i < x$ означає, що для визначення функції розподілу при заданій змінній x необхідно додавати частоти всіх інтервалів, у яких ліва границя менша від значення x. Для побудови графіку функції $F^*(x)$ на осі абсцис позначають границі інтервалів, на осі ординат – відповідні їм значення функції, а отримані точки з'єднують відрізками прямих.

При описуванні окремих ознак розподілу обчислюють відповідні їм числові характеристики. Розглянуті раніше математичні характеристики визначають і для статистичних розподілів. Кожна числова характеристика випадкової величини X має свій статистичний аналог.

Для математичного сподівання аналогом є середнє арифметичне спостережених значень випадкової величини (його ще називають середнім статистичним):
$$M^*[X] = m_x^* = \frac{\sum_{i=1}^{n} x_i}{n}, \tag{1.1.4}$$
де x_i – значення величини, отримане в i-му випробуванні; n – число випробувань. Відповідно до закону великих чисел при $n \to \infty$ середнє арифметичне прямує за ймовірністю до математичного сподівання випадкової величини. При достатньо великому n середнє арифметичне можна вважати приблизно рівним математичному сподіванню.

Аналогом для дисперсії є статистична дисперсія:

$$D_x^* = \frac{\sum_{i=1}^{n}(x_i - m_x^*)^2}{n}. \qquad (1.1.5)$$

Статистичний розподіл має статистичні початкові та центральні моменти різних порядків:

$$\alpha_s^* = \frac{\sum_{i=1}^{n} x_i^s}{n}; \qquad (1.1.6)$$

$$\mu_s^* = \frac{\sum_{i=1}^{n}(x_i - m_x^*)^s}{n}. \qquad (1.1.7)$$

Формулювання статистичних характеристик аналогічні формулюванням числових характеристик випадкової величини. При збільшенні n всі статистичні числові характеристики прямують за ймовірністю до відповідних математичних характеристик і вважаються приблизно рівними їм.

При великому числі випробувань розрахунок характеристик на основі простої статистичної сукупності стає громіздким і незручним. В таких випадках можна використати статистичний ряду розподілу, обчислюючи характеристики за відносними частотами попадання значень величини в інтервали ряду. Значення, яких набуває випадкова величина в i-му інтервалі статистичного ряду можна вважати практично постійними. Тому представником цього інтервалу приймають приблизне значення величини $\tilde{x}_i$, яке дорівнює середньому значенню величини в інтервалі чи навіть середині інтервалу. Тоді статистичні числові характеристики виражаються приблизними формулами такого вигляду:

$$m_x^* = \sum_{i=1}^{k} \tilde{x}_i p_i^*; \qquad (1.1.8)$$

$$D_x^* = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^2 p_i^*; \qquad (1.1.9)$$

$$\alpha_s^* = \sum_{i=1}^{k} \tilde{x}_i^s p_i^*; \qquad (1.1.10)$$

$$\mu_s^* = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^s p_i^*, \qquad (1.1.11)$$

де $\tilde{x}_i$ - представник i-го інтервалу; p_i^* - відносна частота попадання значень величини до відповідного інтервалу; k – число інтервалів ряду.

Статистичне середнє квадратичне відхилення (стандарт) обчислюється за формулою

$$\sigma^* = \sqrt{D_x^*} \,. \tag{1.1.12}$$

Аналітичний вираз функції щільності теоретичного закону розподілу залежить від кількох параметрів - числових характеристик розподілу. При вирівнюванні статистичного ряду розподілу з допомогою теоретичного закону розподілу методом моментів необхідно підібрати для теоретичного закону такі значення характеристик, при яких відповідність між теоретичним і статистичним розподілами буде найкращою. Практично при вирівнюванні статистичного розподілу з допомогою, наприклад, нормального закону, параметрам m_x і σ присвоюють відповідні аналоги m_x^* і σ^* статистичного розподілу. Значення апроксимуючої функції щільності *f(x)* знаходять за умови, що $m_x = m_x^*$ і $\sigma = \sigma^*$. Вирівнюючу криву, яка відповідає функції *f(x)*, для наочності показують на одному малюнку з гістограмою. Отримані таким способом результати разом з їх графічним представленням на вказаному малюнку дають підстави висунути гіпотезу про підпорядкованість статистичного розподілу вибраному теоретичному законові. При наявності розбіжностей між кривою та гістограмою виникає питання: чи ці розбіжності пов'язані з обмеженим числом випробувань над випадковою величиною, чи вони є наслідком невдалого вибору теоретичної вирівнюючої кривої? Відповідь на це питання дають критерії перевірки правдоподібності гіпотез, чи, по-іншому, перевірки узгодженості теоретичного і статистичного розподілів.

Критерій χ^2 *Пірсона*. При розв'язуванні поставлених завдань критерій Пірсона порівняно з іншими критеріями застосовується найчастіше і вважається найбільш строгим з імовірнісної точки зору.

Нехай висунуто гіпотезу про підпорядкованість випадкової величини X певному теоретичному законові розподілу. Знаючи такий закон можна обчислити теоретичні ймовірності попадання випадкової величини в кожний з інтервалів статистичного ряду розподілу. Для перевірки узгодженості розподілів використовуються розбіжності U між теоретичними ймовірностями p_i попадання величини в інтервали та відносними частотами інтервалів p_i^* (i=1,2,...,k; k – число інтервалів статистичного ряду). Мірою розбіжностей U є величина

$$\chi^2 = n\sum_{i=1}^{k} \frac{(p_i^* - p_i)^2}{p_i} \tag{1.1.13}$$

або, враховуючи формулу (1.1.1),

$$U = \chi^2 = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}, \qquad (1.1.14)$$

де m_i - число значень величини в інтервалі; *n* – загальне число випробувань.

Закон розподілу міри *U* наближається до так званого "розподілу χ^2". Розподілом χ^2 з *r* ступенями свободи називається розподіл суми квадратів *r* незалежних випадкових величин, кожна з яких підпорядковується нормальному законові з математичним сподіванням $m_x = 0$ та дисперсією $D_x = 1$. Число ступенів свободи $r = k - s$, (1.1.15)
де *s* – число незалежних умов, накладених на статистичний розподіл. Наприклад, при перевірці підпорядкованості випадкової величини нормальному законові розподілу, *s* = 3, а умовами, які накладені на розподіл, є наступні:

1) $\sum_{i=1}^{k} p_i^* = 1$; 2) $m_x^* = m_x$; 3) $D_x^* = D_x$.

Для розподілу χ^2 складено спеціальну таблицю (див. додаток 1). Користуючись таблицею, кожному значенню χ^2 при потрібному числі ступенів свободи *r* можна визначити ймовірність *p* того, що міра *U*, розподілена за законом χ^2, буде не меншою свого обчисленого значення. Отже, розподіл χ^2 дає можливість кількісно оцінити ступінь розбіжностей теоретичного і статистичного розподілів і виконати перевірку правдоподібності висунутої гіпотези.

Критерій А.Н.Колмогорова. Мірою розбіжностей між теоретичним і статистичним розподілами в цьому критерії є максимальне значення модуля різниці між статистичною функцією розподілу $F^*(x)$ і функцією розподілу $F(x)$ теоретичного закону:

$$D = \max\left|F^*(x) - F(x)\right|. \qquad (1.1.16)$$

Величина *D* просто обчислюється і має досить простий закон розподілу. Доведено, що якою б не була функція розподілу $F(x)$ неперервної випадкової величини *X*, при необмеженому зростанні числа незалежних випробувань *n* ймовірність нерівності $D\sqrt{n} \geq \lambda$ прямує до границі

$$P(\lambda) = 1 - \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2}. \qquad (1.1.17)$$

Для ймовірностей (1.1.17) складено таблицю 1:

Таблиця 1

Значення ймовірностей *P(λ)*.

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

Ймовірність *P(λ)* - це ймовірність того, що за умови підпорядкованості випадкової величини *X* законові *F(x)* внаслідок чисто випадкових причин максимальна розбіжність *D* між $F^*(x)$ та $F(x)$ буде не меншою, ніж фактично спостережена. Якщо ймовірність *P(λ)* мала, то гіпотеза про узгодженість теоретичного і статистичного розподілів відкидається як хибна; при великих *P(λ)* гіпотеза сумісна з дослідними даними.

Критерій А.Н.Колмогорова досить простий, однак його можна використовувати лише в 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 всі параметри, які входять до її складу. Такі випадки на практиці трапляються рідко. Як правило, з певних теоретичних міркувань відомо лише загальний вигляд функції *F(x)*, а її числові параметри визначаються на основі даного статистичного матеріалу (як це роблять, наприклад, використовуючи метод моментів). Критерій Пірсона таку ваду усуває зменшенням числа ступенів свободи розподілу χ^2. Тому, застосовуючи критерій А.Н.Колмогорова при вирівнюванні статистичних рядів методом моментів, одержуємо здебільшого завищене значення ймовірності *P(λ)* і можемо прийняти правдоподібною таку гіпотезу, яка фактично погано узгоджується з дослідними даними.

Якщо статистичний розподіл випадкової величини вирівнюється з допомогою нормального закону і умова задачі дозволяє приблизну перевірку відповідної гіпотези, то в такому випадку достатньо використати досить простий *приблизний критерій*.

Відповідно до 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:

$$S_k^* = \frac{\mu_3^*}{\sigma^{*3}}; \qquad (1.1.18)$$

$$E_x^* = \frac{\mu_4^*}{\sigma^{*4}} - 3, \qquad (1.1.19)$$

де μ_3^* і μ_4^* - відповідно третій та четвертий статистичні центральні моменти; σ^* - статистичне середнє квадратичне відхилення (стандарт). Доведено, що гіпотезу нормального розподілу випадкової величини приблизно можна вважати сумісною з дослідними даними, якщо відхилення від нуля статистичних коефіцієнту асиметрії та ексцесу не перевищують певних допустимих значень, тобто якщо одночасно виконуються умови нерівностей наступного вигляду:

$$\left.\begin{aligned} \left|S_k^*\right| &\le 3\sqrt{\frac{6(n-1)}{(n+1)(n+3)}} \\ \left|E_x^*\right| &\le 5\sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}} \end{aligned}\right\}. \qquad (1.1.20)$$

Такий критерій вигідно виділяється серед інших простотою свого практичного використання і забезпечує приблизне розв'язування завдання перевірки узгодженості розподілу випадкової величини з нормальним законом розподілу.

Порядок виконання роботи

Під час виконання роботи проводиться статистична обробка і аналіз результатів незалежних випробувань, проведених над випадковою величиною, з метою вирівнювання статистичного розподілу величини за допомогою нормального закону і перевірки гіпотези про підпорядкування величини цьому закону розподілу.

Завдання 1. Задано статистичний ряд розподілу похибок вимірювання кутів тріангуляції. Необхідно вирівняти даний експериментальний матеріал з допомогою нормального закону і перевірити гіпотезу про підпорядкованість похибок вимірів кутів цьому теоретичному законові розподілу випадкової величини.

Вихідні дані для виконання завдання наведено в таблиці додатку 4. Дані слід вибрати відповідно до номера варіанту N студента.

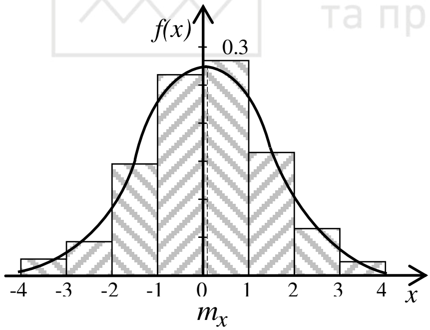
Приклад. При n=527 результатах випробувань їх статистичну обробку зручніше проводити на основі статистичного ряду розподілу похибок. Тому на першому етапі обробки перетворюємо просту статистичну сукупність похибок вимірів кутів до статистичного ряду розподілу похибок.

1. *Побудова статистичного ряду розподілу похибок вимірів кутів.* Експериментальні дані показують, що значення похибок вимірів кутів змінюються в діапазоні від -4 до +4. Для зручності розділимо цей діапазон на k=8 інтервалів, підрахуємо число попадань помилок в межі кожного інтервалу m_i і за формулою (1.1.1) обчислимо відносні частоти (статистичні ймовірності) p_i^* попадання похибок до кожного інтервалу ($i = 1,2,...,8$). Отримані результати запишемо в таблицю – статистичний ряд розподілу похибок вимірів кутів:

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
p_i^*	0.021	0.045	0.148	0.264	0.281	0.161	0.061	0.019

Контроль обчислень: $\sum_{i=1}^{k} m_i = n = 527$; $\sum_{i=1}^{k} p_i^* = 1.000$.

2. *Графічне зображення статистичного ряду розподілу помилок вимірів кутів у вигляді гістограми*. Для побудови гістограми на осі абсцис позначуємо інтервали і на кожному з них, як на основі, будуємо прямокутник. Площа прямокутника повинна дорівнювати частоті відповідного інтервалу. Повна площа гістограми дорівнює одиниці (див. малюнок 1.1.1).



Малюнок 1.1.1

Взаємне розташування прямокутників гістограми за своїм виглядом подібне формі кривої розподілу нормального закону. Тому, виходячи із зовнішнього вигляду гістограми як графічного зображення статистичного ряду розподілу помилок вимірів кутів, є підстави вирівнювання такого ряду виконувати кривою нормального закону розподілу випадкової величини. Вирівнювання виконуємо із застосуванням методу моментів, який передбачає, що теоретичним значенням числових характеристик математичного сподівання та стандарту, які входять до аналітичного виразу функції щільності нормального закону, присвоюються відповідні їм статистичні аналоги.

3. *Обчислення статистичних характеристик розподілу похибок* (математичного сподівання m_x^* і стандарту σ^*).

Статистичні значення числових характеристик m_x^* та σ^* визначаємо за формулами (1.1.8), (1.1.9) та (1.1.12). Результати проміжних розрахунків показані 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

$\tilde{x}_i$ - середина кожного i-го інтервалу статистичного ряду розподілу.

Отже, математичне сподівання (або статистичне середнє помилок вимірів кутів) $m_x^* = \sum_{i=1}^{k} \tilde{x}_i p_i^* = 0.060.$ m_x^* характеризує положення центру статистичного розподілу помилок.

Характеристики розсіювання статистичного розподілу похибок: статистична дисперсія $D_x^* = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^2 p_i^* = 1.9802;$ статистичне середнє квадратичне відхилення (стандарт) $\sigma^* = \sqrt{D_x^*} = 1.407.$

4. *Побудова вирівнюючої кривої нормального закону розподілу*. Для побудови вирівнюючої кривої скористаємось аналітичним виразом функції щільності нормального закону і обчислимо значення функції для середини інтервалів $\tilde{x}_i$, покладаючи, згідно методу моментів, що $m_x = m_x^*$ та $\sigma = \sigma^*$. Використаємо таблиці функції $f(t_i)$ (див. додаток 1 метод. вказівок **05-04-30**). Нормоване значення величини $t_i = \dfrac{\tilde{x}_i - m_x}{\sigma}$. Кінцеві значення функції $f(\tilde{x}_i) = \dfrac{f(t_i)}{\sigma}$. Результати обчислень показані 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

Якщо довжини інтервалів не дорівнюють одиниці, то для знаходження кінцевих значень функції щільності обчислені в останній таблиці значення $f(\tilde{x}_i)$ слід множити на довжини інтервалів, яким ті відповідають.

Оскільки площа під кривою розподілу дорівнює одиниці, то сума значень функції щільності для всіх інтервалів також повинна дорівнювати одиниці. Звідси слідує контроль обчислення $f(\tilde{x}_i)$: $\sum_{i=1}^{k} f(\tilde{x}_i) = 0.996 \approx 1.$

За значеннями $f(\tilde{x}_i)$ на одному малюнку з гістограмою будуємо вирівнюючу криву, яка відповідає нормальному законові розподілу (див. малюнок 1.1.1). Виходячи з того, що крива добре описує гістограму, можна висунути гіпотезу, що помилки вимірів кутів підпорядковуються нормальному законові розподілу.

5. *Перевірка правдоподібності гіпотези за критерієм* χ^2 *Пірсона*:

а) обчислення ймовірностей попадання нормально розподіленої випадкової величини в інтервали статистичного ряду розподілу за формулою:

$$p_i = \Phi^*(t_{i_{пр}}) - \Phi^*(t_{i_л}) = \Phi^*\left(\frac{x_{i_{пр}} - m_x^*}{\sigma^*}\right) - \Phi^*\left(\frac{x_{i_л} - m_x^*}{\sigma^*}\right).$$

$x_{i_{пр}}$ та $x_{i_л}$ - права та ліва границі кожного інтервалу. Значення $\Phi^*(t)$ визначаємо з таблиці додатку 2 метод. вказівок **05-04-30**. Результати обчислення теоретичних ймовірностей p_i :

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

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

б) обчислення міри розбіжностей статистичного і теоретичного розподілів $U = \chi^2$ за формулою (1.1.14):

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

Отже, $\chi^2 = 5.098$.

в) обчислення числа ступенів свободи r. Оскільки при вирівнюванні статистичних рядів з допомогою нормального закону методом моментів число накладених умов $s = 3$, то $r = k - s = 5$.

г) визначення ймовірності $p(\chi^2)$ за таблицею додатку 3. При $\chi^2 = 5.098$ та $r = 5$ ймовірність $p(\chi^2) = 0.42$.

Ймовірність $p(\chi^2) = 0.42$ того, що міра розбіжностей χ^2 не менша свого обчисленого значення, є досить значною, тобто існуючі розбіжності між теоретичним і статистичним розподілами викликані випадковими причинами, пов'язаними з обмеженим числом випробувань. При $0.3 \le p(\chi^2) < 0.5$ узгодженість розподілів добра. Отже, гіпотезу про підпорядкованість помилок вимірів кутів нормальному законові розподілу можна прийняти і вважати такою, що не заперечує дослідних даних.

6.*Перевірка правдоподібності гіпотези за критерієм А.Н.Колмогорова*.

а) побудова статистичної функції розподілу помилок вимірів кутів. Значення статистичної функції розподілу $F^*(x)$ визначаємо за формулою (1.1.3) на основі даних статистичного ряду розподілу помилок:

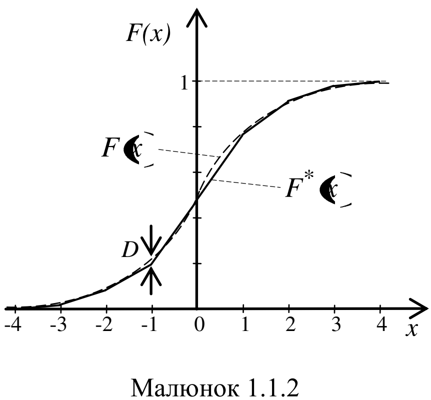
$F^*(-4) = 0$; $F^*(-3) = p_1^* = 0.021$; $F^*(-2) = p_1^* + p_2^* = 0.066$;

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

$F^*(1) = p_1^* + p_2^* + \ldots + p_5^* = 0.759$; $F^*(2) = p_1^* + p_2^* + \ldots + p_6^* = 0.920$;

$F^*(3) = p_1^* + p_2^* + \ldots + p_7^* = 0.981$; $F^*(4) = p_1^* + p_2^* + \ldots + p_8^* = 1.000$.

За обчисленими значеннями будуємо графік статистичної функції розподілу $F^*(x)$ помилок вимірів кутів (див. малюнок 1.1.2).



Малюнок 1.1.2

б) побудова теоретичної (нормальної) функції розподілу помилок вимірів кутів. Значення нормальної функції розподілу $F(x_i) = \Phi^*(x_i)$ при x_i, які дорівнюють границям інтервалів статистичного ряду, обчислені в п.5.а та показані у відповідній таблиці. За цими значеннями будуємо графік теоретичної функції розподілу помилок вимірів кутів (див. малюнок 1.1.2).

Результати обчислення значень статистичної та теоретичної функцій розподілу зведемо до таблиці:

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

в) оцінка узгодженості статистичного та теоретичного розподілів. З останньої таблиці та малюнку 1.1.2 видно, що величина міри розбіжностей критерію Колмогорова, яка виражається максимальною різницею значень статистичної та теоретичної функцій розподілу, відповідає значенню $x = -1$: $D = \max\left\lvert F^*(x_i) - F(x_i) \right\rvert = 0.012$. Обчислюємо величину $\lambda = D\sqrt{n} = 0.28$ і за таблицею додатку 4 визначаємо ймовірність *P(λ)* того, що за умови підпорядкованості помилок вимірів кутів законові розподілу $F(x)$ внаслідок чисто випадкових причин максимальна розбіжність *D* між $F^*(x)$ та $F(x)$ буде не меншою, ніж фактично обчислена. Отже, *P(λ)*= 1.000. Подію з такою ймовірністю появи вважають достовірною. Тому узгодженість статистичного і теоретичного розподілів відмінна, а гіпотезу про підпорядкованість помилок вимірів кутів нормальному законові розподілу можна вважати сумісною з дослідними даними.

7. *Перевірка правдоподібності гіпотези за приблизним критерієм.* Обчислюємо статистичні значення коефіцієнту асиметрії S_k^* та ексцесу E_x^*. Для цього можна скористатись формулами (1.1.18) та (1.1.19). Результати проміжних обчислень показані в таблиці:

$\tilde{x}_i$	-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

$$\mu_3^* = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^3 p_i^* = -0.146 \quad \text{і} \quad S_k^* = -0.052;$$

$$\mu_4^* = \sum_{i=1}^{k} (\tilde{x}_i - m_x^*)^4 p_i^* = 11.733 \quad \text{і} \quad E_x^* = -0.006.$$

Отже, характеристики асиметрії та ексцесу не дорівнюють нулю. Асиметрія статистичного розподілу помилок від'ємна, тобто крива розподілу зміщена вліво порівняно з кривою нормального закону. Ексцес статистичного розподілу незначний. Його величина вказує на те, що крива цього розподілу є більш плоско вершинною порівняно з кривою нормального розподілу.

Визначимо допустимі відхилення від нуля коефіцієнта асиметрії та ексцесу і перевіримо умови нерівностей (1.1.20):

$$3\sqrt{\frac{6\times 526}{528\times 530}} = 0.315; \qquad 5\sqrt{\frac{24\times 527\times 525\times 524}{528^2\times 530\times 532}} = 1.049;$$

$$\left|S_k^*\right| < 0.315; \qquad \left|E_x^*\right| < 1.049.$$

Умови обох нерівностей виконуються, відхилення від нуля коефіцієнта асиметрії та ексцесу не перевищують відповідних їм допустимих значень. Тому, незважаючи на те, що характеристики асиметрії і ексцесу не дорівнюють нулю, гіпотеза про підпорядкованість помилок вимірів кутів нормальному законові може бути прийнята як правдоподібна.

Питання для самостійної роботи

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

2. ВИЗНАЧЕННЯ ТОЧНОСТІ ТА НАДІЙНОСТІ ЧИСЛОВИХ ОЦІНОК ПАРАМЕТРІВ РОЗПОДІЛУ ПРИ ОБМЕЖЕНОМУ ЧИСЛІ ВИПРОБУВАНЬ

Основні теоретичні положення

На практиці дуже часто статистичний матеріал обмежений кількома десятками випробувань, тому на його основі неможливо встановити закон розподілу величини. Однак його можна використати для визначення певних відомостей про випадкову величину, наприклад, обчислити її числові характеристики. Трапляються випадки, коли закон розподілу величини відомий, а необхідно знайти лише параметри, від яких він залежить. Так, якщо відомо, що величина розподіляється за нормальним законом, то завдання обробки випробувань зводиться до визначення двох основних його параметрів – математичного сподівання та стандарту. В окремих задачах вид закону розподілу взагалі не суттєвий, а необхідно знати лише його числові характеристики. Отже, завдання обробки обмеженого числа випробувань зводиться до визначення невідомих параметрів – числових характеристик, від яких залежить закон розподілу величини.

Приблизне значення $\tilde{a}$ невідомого параметру *a*, яке обчислене на основі обмеженого числа випробувань, називають оцінкою цього параметру. Оцінка $\tilde{a}$ деякого параметру *a* є функцією результатів випробувань $x_1, x_2, ..., x_n$ над випадковою величиною *X* вигляду $\tilde{a} = a(x_1, x_2, ..., x_n)$. Тому оцінку $\tilde{a}$ слід також розглядати як випадкову величину. Будь-яка оцінка $\tilde{a}$ параметру *a* повинна задовольняти наступним вимогам:

1)оцінка параметру повинна бути найбільш надійною, тобто такою, щоб при збільшенні числа випробувань вона прямувала за ймовірністю до самого параметру, тобто $\tilde{a} \to a$ при $n \to \infty$;

2)оцінка параметру мусить бути не зміщеною - вона не повинна включати систематичної помилки і її математичне сподівання має дорівнювати самому параметрові, тобто $M[\tilde{a}] = a$;

3)оцінка мусить бути ефективною, для чого її дисперсія має бути мінімальною: $D[\tilde{a}] = \min$.

Оцінка $\tilde{a}$ невідомого параметру *a* розкривається окремим числом, тому її називають точковою (або числовою) оцінкою параметру. В деяких задачах необхідно не лише описати параметр *a* числовим значенням $\tilde{a}$, але й охарактеризувати точність та надійність цього значення. Це найбільш важливо при малому числі випробувань, коли оцінка $\tilde{a}$ в значній мірі випадкова і наближена заміна *a* на $\tilde{a}$ може викликати серйозні помилки. Щоб вказати точність та надійність оцінки $\tilde{a}$, в математичній статистиці користуються поняттями довірчого інтервалу та довірчої ймовірності.

Установимо певну досить велику ймовірність β таку, що подію з такою ймовірністю можна вважати практично достовірною. Знайдемо значення помилки ε, для якого: $$P(\tilde{a}-\varepsilon<a<\tilde{a}+\varepsilon)=\beta. \quad (1.2.1)$$

Тоді діапазон практично можливих значень помилки заміни a на $\tilde{a}$ дорівнюватиме $\pm\varepsilon$. Рівність (1.2.1) означає, що з ймовірністю β невідоме значення параметру a потрапить в інтервал : $$I_\beta=(\tilde{a}-\varepsilon;\tilde{a}+\varepsilon), \quad (1.2.2)$$

або, по-іншому, β - це ймовірність того, що випадковий інтервал I_β покриє точку a .

Ймовірність β називають довірчою ймовірністю і використовують для судження про надійність числової оцінки $\tilde{a}$. Інтервал I_β називають довірчим інтервалом і розглядають як характеристику точності числової оцінки $\tilde{a}$ параметру a. Границі довірчого інтервалу $a_1=\tilde{a}-\varepsilon$ та $a_2=\tilde{a}+\varepsilon$ називають довірчими границями. Довірчий інтервал слід розуміти як інтервал можливих значень невідомого параметру a, сумісних з дослідними даними і які не суперечать їм.

Розглянемо правила формулювання числових оцінок та визначення їх точності та надійності на прикладі основних параметрів розподілу випадкової величини – математичного сподівання та дисперсії (чи стандарту).

Якщо число n проведених випробувань велике, то знаходять середнє статистичне значення випадкової величини (статистичний аналог математичного сподівання) за формулою (1.1.4) та статистичні дисперсію чи стандарт за формулами (1.1.5) чи (1.1.12). При обмеженому числі випробувань можна визначити лише приблизні значення цих невідомих параметрів – числові оцінки математичного сподівання та дисперсії (стандарту).

Числовою оцінкою $\tilde{m}_x$ математичного сподівання є статистичне середнє результатів випробувань:

$$\tilde{m}_x=m_x^*=\frac{\sum_{i=1}^{n}x_i}{n} . \quad (1.2.3)$$

Числовою оцінкою $\tilde{D}_x$ дисперсії приймають величину

$$\tilde{D}_x=\frac{n}{n-1}D_x^*=\frac{\sum_{i=1}^{n}(x_i-\tilde{m}_x)^2}{n-1} . \quad (1.2.4)$$

Доведено, що числові оцінки $\tilde{m}_x$ та $\tilde{D}_x$ задовольняють всім вимогам оцінок параметрів, які обчислені за обмеженим числом випробувань, тобто вони є найбільш надійними, не зміщеними та ефективними.

При описуванні точності та надійності $\tilde{m}_x$ та $\tilde{D}_x$ в залежності від формулювання розв'язуваної задачі та виду закону розподілу випадкової величини користуються *приблизними* або *точними* методами побудови довірчих інтервалів.

Приблизні методи побудови довірчих інтервалів для математичного сподівання чи дисперсії використовують в тих 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 оцінок параметрів. Оцінка $\tilde{m}_x$ є сумою n незалежних однаково розподілених значень випадкової величини x_i, тому, відповідно до центральної граничної теореми, можна вважати, що вона розподілена за нормальним законом. Навіть при числі випробувань порядку 10÷20 закон розподілу суми приблизно вважається нормальним. Виходячи з цього, можна визначити значення помилки ε_β при вибраній довірчій імовірності β, для якої $P(\tilde{m}_x - \varepsilon_\beta < m_x < \tilde{m}_x + \varepsilon_\beta) = \beta$.

Така ймовірність виражається через нормальну функцію розподілу $2\Phi^*\left(\frac{\varepsilon_\beta}{\sigma_{\tilde{m}}}\right) - 1 = \beta$, звідки:

$$\varepsilon_\beta = \sigma_{\tilde{m}} \arg\Phi^*\left(\frac{1+\beta}{2}\right), \tag{1.2.5}$$

де $\arg\Phi^*\left(\frac{1+\beta}{2}\right)$ - функція, обернена до $\Phi^*\left(\frac{1+\beta}{2}\right)$, тобто таке значення аргументу, при якому нормальна функція розподілу дорівнює $\frac{1+\beta}{2}$; $\sigma_{\tilde{m}}$ - стандарт оцінки $\tilde{m}_x$, наближене значення якого можна обчислити за формулою

$$\sigma_{\tilde{m}} = \sqrt{\frac{\tilde{D}_x}{n}}. \tag{1.2.6}$$

Отже, при вибраній довірчій імовірності β довірчий інтервал для математичного сподівання окреслюватиметься оцінкою $\tilde{m}_x$ та помилкою ε_β і матиме вигляд

$$I_\beta = (\tilde{m}_x - \varepsilon_\beta; \tilde{m}_x + \varepsilon_\beta). \tag{1.2.7}$$

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 $\tilde{m}_x$. Закон розподілу такої суми також гранично наближається до нормального. Практично, така гіпотеза може бути прийнятою вже при числі випробувань n = 20÷30. На цій підставі довірчий інтервал для дисперсії, що відповідає довірчій імовірності $\beta = P(\tilde{D}_x - \varepsilon_\beta < D_x < \tilde{D}_x + \varepsilon_\beta)$, окреслюватиметься оцінкою $\tilde{D}_x$ та відповідною помилкою ε_β і матиме вигляд

$$I_\beta = (\tilde{D}_x - \varepsilon_\beta; \tilde{D}_x + \varepsilon_\beta), \tag{1.2.8}$$

де помилка

$$\varepsilon_\beta = \sigma_{\tilde{D}} \arg \Phi^*\left(\frac{1+\beta}{2}\right), \tag{1.2.9}$$

а стандарт $\sigma_{\tilde{D}}$ оцінки $\tilde{D}_x$ приблизно можна підрахувати за формулою

$$\sigma_{\tilde{D}} = \tilde{D}_x \sqrt{\frac{2}{n-1}}\,. \tag{1.2.10}$$

Якщо є підстави вважати, що закон розподілу випадкової величини суттєво відрізняється від нормального і розбіжності носять принциповий характер, то значення $\sigma_{\tilde{D}}$ розраховують по-іншому. Наприклад, якщо має місце рівномірний закон розподілу величини X, то $\sigma_{\tilde{D}} \approx \tilde{D}_x \sqrt{\frac{0.8n+1.2}{n(n-1)}}$. Однак при невідомому законові розподілу величини значення $\sigma_{\tilde{D}}$ рекомендується приблизно установлювати за формулою (1.2.10).

Точні методи побудови довірчих інтервалів для невідомих параметрів розподілу використовують у випадках, коли заздалегідь відомо, що випадкова величина розподіляється за нормальним законом.

Доведено, що при нормальному розподілі величини X випадкова величина $T = \sqrt{n}\,\frac{\tilde{m}_x - m_x}{\sqrt{\tilde{D}_x}}$ підпорядковується так званому законові розподілу Стьюдента з $n-1$ ступенями свободи, опираючись на який при заданій довірчій імовірності β будують довірчий інтервал виду (1.2.7) для математичного сподівання випадкової величини: $I_\beta = (\tilde{m}_x - \varepsilon_\beta; \tilde{m}_x + \varepsilon_\beta)$. Значення помилки ε_β знаходять з формули

$$\varepsilon_\beta = t_\beta \sqrt{\frac{\tilde{D}_x}{n}} \tag{1.2.11}$$

за умови $P(|\tilde{m}_x - m_x| < \varepsilon_\beta) = \beta$. Параметр t_β визначається за ймовірністю β з інтегралу

$$2\int_0^{t_\beta} S_{n-1}(t)\,dt = \beta \tag{1.2.12}$$

за умови $P(|T| < t_\beta) = \beta$, де $S_{n-1}(t)$ - щільність закону розподілу Стьюдента з $n-1$ ступенями свободи. Для інтегралу (1.2.12) складено таблиці, користуючись якими, за значенням довірчої ймовірності β і числом ступенів свободи $n-1$ визначається параметр t_β (див. таблицю додатку).

Доведено також, що випадкова величина $V = \frac{(n-1)\tilde{D}_x}{D_x}$ при нормальному розподілі величини X має розподіл χ^2 з $r = n-1$ ступенями свободи, опираючись на який при заданій довірчій імовірності β будують довірчий інтервал для дисперсії. Випадкова величина $\tilde{D}_x$ виражається через величину V, яка має розподіл χ^2, залежністю $\tilde{D}_x = V\frac{D_x}{n-1}$.

Знаючи закон розподілу величини V, можна знайти інтервал $i_\beta = (\chi_1^2; \chi_2^2)$, в який вона попадає із заданою ймовірністю β (χ_1^2 та χ_2^2 - границі інтервалу). Ймовірність попадання величини V поза межі інтервалу i_β буде $\alpha = 1 - \beta$. Щоб побудувати інтервал i_β, використовують таблицю розподілу χ^2 (див. таблицю додатку 3). В ній наведено значення χ^2 такі, що $P(V > \chi^2) = p$ для величини V, яка має розподіл χ^2 з $r = n-1$ ступенями свободи. 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 $p_2 = 1 - \frac{\alpha}{2}$.

Потрібний нам довірчий інтервал для дисперсії $I_\beta = (D_1; D_2)$ з границями D_1 і D_2, який покриває невідоме значення D_x з ймовірністю $\beta = P(D_1 < D_x < D_2)$, покриває його тоді і тільки тоді, коли величина V потрапляє в інтервал i_β. Тому інтервал I_β для дисперсії можна знайти за інтервалом i_β для величини V. Границі D_1 та D_2 довірчого інтервалу I_β для дисперсії при заданій довірчій імовірності β знаходять за формулами:

$$D_1 = \frac{\tilde{D}_x (n-1)}{\chi_1^2}; \qquad D_2 = \frac{\tilde{D}_x (n-1)}{\chi_2^2}. \tag{1.2.13}$$

Порядок виконання роботи

Під час виконання роботи рахуються числові оцінки математичного сподівання та дисперсії, а також визначаються їх точність і надійність приблизними та точними методами. На практиці при побудові довірчих інтервалів залежно від поставленого завдання користуються або приблизними, або точними методами. Використання ж в даній роботі одночасно і точних, і приблизних методів пов'язане лише з необхідністю освоїти правила і порядок практичного використання цих методів.

Завдання 2. Задано статистичний ряд розподілу похибок вимірів кутів тріангуляції. Необхідно визначити оцінки математичного сподівання $\tilde{m}_x$ та дисперсії $\tilde{D}_x$ для цього розподілу, побудувати довірчі інтервали оцінок $\tilde{m}_x$ та $\tilde{D}_x$ приблизними і точними методами при довірчій імовірності β=0.9. Числові значення вихідних даних слід прийняти такими ж, як і в практичній роботі №1 «Визначення закону розподілу на основі дослідних даних».

Приклад. Дано статистичні значення математичного сподівання m_x^*=0.060 та дисперсії D_x^*=1.9802 розподілу похибок вимірів кутів тріангуляції, які обчислено з n=20 результатів вимірів.

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

Відповідно до формули (1.2.3) оцінка математичного сподівання $\tilde{m}_x = m_x^* = 0.060$. Оцінку дисперсії визначимо за формулою (1.2.4): $\tilde{D}_x = \dfrac{n}{n-1} D_x^* = 2.0844$.

2. *Побудова довірчих інтервалів приблизними методами.*

а) для математичного сподівання:

Довірчий інтервал $I_\beta = (\tilde{m}_x - \varepsilon_\beta; \tilde{m}_x + \varepsilon_\beta)$ при довірчій імовірності β=0.9 для математичного сподівання будуємо застосовуванням формул (1.2.5) та (1.2.6): $\sigma_{\tilde{m}} = \sqrt{\dfrac{\tilde{D}_x}{n}} = 0.3228$; $\varepsilon_\beta = \sigma_{\tilde{m}} \arg \Phi^* \left(\dfrac{1+\beta}{2} \right) = 0.531$.

Остаточно I_β=(-0.471; 0.591).

б) для дисперсії:

Довірчий інтервал $I_\beta = (\tilde{D}_x - \varepsilon_\beta; \tilde{D}_x + \varepsilon_\beta)$ при довірчій імовірності β=0.9 для дисперсії будуємо застосовуванням формул (1.2.9) та (1.2.10): $\sigma_{\tilde{D}} = \tilde{D}_x \sqrt{\dfrac{2}{n-1}} = 0.6763$; $\varepsilon_\beta = \sigma_{\tilde{D}} \arg \Phi^* \left(\dfrac{1+\beta}{2} \right) = 1.1125$. Остаточно I_β=(0.9719; 3.1969).

3. *Побудова довірчих інтервалів точними методами.*

а) для математичного сподівання:

Довірчий інтервал $I_\beta = (\tilde{m}_x - \varepsilon_\beta; \tilde{m}_x + \varepsilon_\beta)$ при довірчій імовірності β=0.9 для математичного сподівання будуємо застосовуванням формули (1.2.11) і таблиці параметру розподілу Стьюдента (див. таблицю додатку 2). З таблиці за числом ступенів свободи n-1=19 та β=0.9 знаходимо значення параметру t_β=1.729. Тоді $\varepsilon_\beta = t_\beta \sqrt{\frac{\tilde{D}_x}{n}}$ =0.558. Остаточно I_β=(-0.498; 0.618).

б) для дисперсії:

Довірчий інтервал $I_\beta = (D_1; D_2)$ при довірчій імовірності β=0.9 для дисперсії будуємо застосовуванням формул (1.2.13) і таблиці розподілу χ^2 (див. таблицю додатку 3) за наступними правилами. З таблиці за числом ступенів свободи r =n-1=19 та ймовірністю $p_1 = \frac{1-\beta}{2} = 0.05$ знаходимо значення χ_1^2=30.1, а за ймовірністю $p_2 = 1 - \frac{1-\beta}{2} = 0.95$ - значення χ_2^2=10.11. Довірчий інтервал $i_\beta = (\chi_1^2; \chi_2^2) = (30.1; 10.11)$. Тоді границі довірчого інтервалу для дисперсії $D_1 = \frac{\tilde{D}_x (n-1)}{\chi_1^2}$=1.3157 та $D_2 = \frac{\tilde{D}_x (n-1)}{\chi_2^2}$= =3.9173. Остаточно I_β= (1.3157; 3.9173).

Питання для самостійної роботи

1. Що називають оцінкою невідомого параметру розподілу випадкової величини?
2. Яким вимогам повинна задовольняти оцінка параметру розподілу?
3. Як визначити числові оцінки математичного сподівання та дисперсії?
4. Що називають довірчою ймовірністю, довірчим інтервалом?
5. В чому полягає завдання оцінки точності та надійності невідомого параметру розподілу?
6. Як оцінити точність та надійність числової оцінки математичного сподівання приблизним методом?
7. Як оцінити точність та надійність числової оцінки дисперсії приблизним методом?
8. Як оцінити точність та надійність числової оцінки математичного сподівання точним методом?
9. Як оцінити точність та надійність числової оцінки дисперсії точним методом?
10. В яких випадках для оцінки точності та надійності невідомих параметрів розподілу випадкової величини користуються приблизними методами? В яких випадках користуються точними методами?

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

Основні теоретичні положення

Поняття залежності в системі випадкових величин має особливий зміст. Це не є строга функціональна математична залежність, коли за значенням однієї величини можна точно вказати значення другої. В теорії ймовірностей розглядається імовірнісна (або стохастична, кореляційна) залежність. Вона передбачає, що, знаючи значення однієї величини, не можна точно вказати значення другої величини, а лише її закон розподілу, який залежить від значень, яких набуває перша випадкова величина.

Властивості системи *(X,Y)* не вичерпуються лише властивостями окремих величин, які входять в систему. Вони включають також взаємні зв'язки між випадковими величинами *X* та *Y*. Встановлення та вираження саме цієї властивості системи випадкових величин являє собою актуальне та важливе з практичної точки зору статистичне завдання. Система дій зі встановлення зв'язку між випадковими величинами системи, а також визначення його тісноти та форми, називається кореляційним аналізом. Кореляційний аналіз передбачає розрахунок низки параметрів, які характеризують систему випадкових величин *(X,Y)*.

Кореляційний момент $K_{x,y}$ системи *(X,Y)* – другий змішаний центральний момент, який виражається формулою

$$K_{x,y} = \sum_i \sum_j (x_i - m_x)(y_j - m_y) p_{ij} . \qquad (1.3.1)$$

Кореляційний момент $K_{x,y}$ є характеристикою розсіювання і зв'язку величин *X* та *У*. Доведено, що для незалежних випадкових величин $K_{x,y} = 0$. Такі величини ще називають некорельованими. Якщо кореляційний момент двох випадкових величин не дорівнює нулю, це є ознакою існування залежності між ними. Якщо одна з величин системи *(X,Y)* досить мало відрізняється від свого математичного сподівання, тобто майже не випадкова, то $K_{x,y}$ буде малим, якою б тісною залежністю не були пов'язані величини. Щоб позбавитись такого недоліку і отримати характеристику зв'язку між величинами *X* та *Y* в чистому вигляді, використовують нормоване значення кореляційного моменту, яке називають коефіцієнтом кореляції.

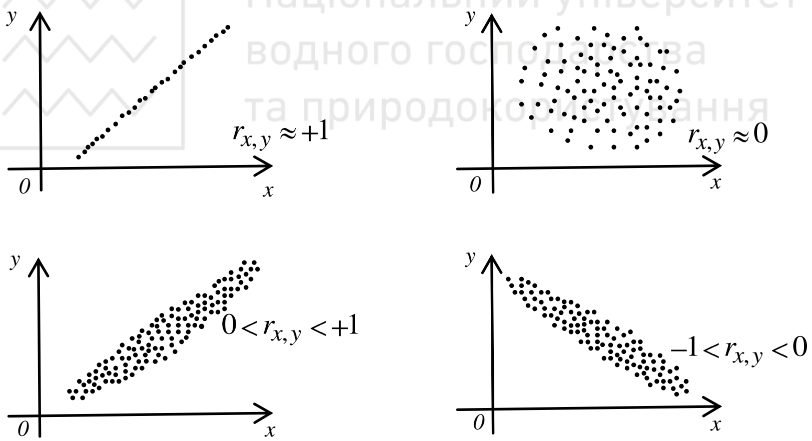
*К*оефіцієнт кореляції $r_{x,y}$ системи *(X,Y)*

$$r_{x,y} = \frac{K_{x,y}}{\sigma_x \sigma_y} . \qquad (1.3.2)$$

Коефіцієнт кореляції $r_{x,y}$ - безрозмірна характеристика ступеню тісноти лінійної імовірнісної залежності між величинами *X* та *Y*.

Лінійна імовірнісна залежність полягає в тому, що при зростанні однієї величини друга має тенденцію зростати чи спадати за лінійним законом. Ця тенденція гранично наближається до функціональної, тобто самої тісної лінійної залежності. Якщо між величинами існує функціональний лінійний зв'язок, то $r_{x,y} = \pm 1$. Загалом, якщо величини X та Y зв'язані довільною імовірнісною залежністю, то $-1 < r_{x,y} < +1$. При $r_{x,y} > 0$ залежність (або кореляція) величин X та Y позитивна, а при $r_{x,y} < 0$ кореляція величин негативна. Чим більшого значення набуває коефіцієнт кореляції, тим тісніший зв'язок між випадковими величинами і тим більше вони залежні між собою. Чим ближче коефіцієнт кореляції до ± 1, тим більше імовірнісна лінійна залежність величин наближається до функціональної залежності.

Якщо відомо результати n повторних випробувань над системою випадкових величин *(X,Y)*, то про можливе існування залежності між величинами можна зробити попередній висновок, виходячи із зовнішнього вигляду діаграми залежності. На такій діаграмі зображують точками всі пари значень випадкових величин X та Y, які отримані в результаті проведених випробувань (див. малюнок 1.1.3).



Малюнок 1.1.3

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 коефіцієнту кореляції

$$r_{x,y}^* = \frac{K_{x,y}^*}{\sigma_x^* \sigma_y^*}, \tag{1.3.3}$$

де статистичні значення кореляційного моменту $K^*_{x,y}$ і середніх квадратичних відхилень (стандартів) σ^*_x та σ^*_y величин X та Y виражають формулами

$$K^*_{x,y}=\frac{\sum\limits_{i=1}^{n}\left(x_i-m^*_x\right)\left(y_i-m^*_y\right)}{n}=\frac{\sum\limits_{i=1}^{n}x_i y_i}{n}-m^*_x m^*_y\,; \qquad (1.3.4)$$

$$\sigma^*_x=\sqrt{D^*_x}\,; \qquad \sigma^*_y=\sqrt{D^*_y}\,. \qquad (1.3.5)$$

Тут m^*_x та m^*_y - статистичні математичні сподівання величин X та Y

$$m^*_x=\frac{\sum\limits_{i=1}^{n}x_i}{n}\,; \qquad m^*_y=\frac{\sum\limits_{i=1}^{n}y_i}{n}\,, \qquad (1.3.6)$$

а D^*_x та D^*_y - статистичні дисперсії величин X та Y

$$D^*_x=\frac{\sum\limits_{i=1}^{n}\left(x_i-m^*_x\right)^2}{n}=\frac{\sum\limits_{i=1}^{n}x_i^2}{n}-\left(m^*_x\right)^2; \qquad D^*_y=\frac{\sum\limits_{i=1}^{n}\left(y_i-m^*_y\right)^2}{n}=\frac{\sum\limits_{i=1}^{n}y_i^2}{n}-\left(m^*_y\right)^2 \qquad (1.3.7)$$

Визначеної за формулою (1.3.3) тісноти достатньо, щоб кореляційний зв'язок вважати існуючим і встановленим, якщо одночасно задовольняються вимоги таких нерівностей:

$$\left.\begin{aligned}\left|r^*_{x,y}\right| &\ge r_{\min}\\ \left|r^*_{x,y}\right| &\ge 3\sigma_r\end{aligned}\right\}, \qquad (1.3.8)$$

де $r_{\min}$ - найменше допустиме значення коефіцієнту кореляції

$$r_{\min}=\frac{\sqrt{n+36}-\sqrt{n}}{6}\,; \qquad (1.3.9)$$

σ_r - середнє квадратичне відхилення коефіцієнту кореляції

$$\sigma_r=\frac{1-r^{*\,2}_{x,y}}{\sqrt{n}}\,. \qquad (1.3.10)$$

Для проведення кореляційного аналізу потрібно мати статистичний матеріал досить великого об'єму – число *n* результатів випробувань повинно бути великим. На практиці такий матеріал переважно кількісно обмежений. Однак ним можна скористатися для отримання певних відомостей про залежність випадкових величин і визначити приблизне значення тісноти кореляційного зв'язку. Приблизне значення коефіцієнту кореляції, обчислене на основі обмеженого числа випробувань, називають оцінкою коефіцієнту кореляції. Оцінка коефіцієнту кореляції і оцінки інших параметрів системи *(X,Y)* повинні бути найбільш надійними, незміщеними і ефективними. Оцінка

$\tilde{r}_{x,y}$ коефіцієнту кореляції $r_{x,y}$ системи випадкових величин *(X,Y)* виражається формулою:

$$\tilde{r}_{x,y} = \frac{\tilde{K}_{x,y}}{\tilde{\sigma}_x \tilde{\sigma}_y}, \qquad (1.3.11)$$

де оцінка кореляційного моменту $\tilde{K}_{x,y}$, оцінки середніх квадратичних відхилень $\tilde{\sigma}_x$ та $\tilde{\sigma}_y$, оцінки дисперсій $\tilde{D}_x$ та $\tilde{D}_y$ і оцінки математичних сподівань $\tilde{m}_x$ та $\tilde{m}_y$ визначаються з таких співвідношень:

$$\tilde{K}_{x,y} = \frac{n}{n-1} K^*_{x,y}; \qquad (1.3.12)$$

$$\tilde{\sigma}_x = \sqrt{\tilde{D}_x}; \qquad \tilde{\sigma}_y = \sqrt{\tilde{D}_y}; \qquad (1.3.13)$$

$$\tilde{D}_x = \frac{n}{n-1} D^*_x; \qquad \tilde{D}_y = \frac{n}{n-1} D^*_y; \qquad (1.3.14)$$

$$\tilde{m}_x = m^*_x; \qquad \tilde{m}_y = m^*_y. \qquad (1.3.15)$$

Оцінка $\tilde{r}_{x,y}$ приблизно описує тісноту зв'язку вже при числі проведених випробувань $n < 50$. Зв'язок між випадковими величинами *X* та *Y* вважається встановленим, якщо одночасно задовольняються умови нерівностей

$$\left.\begin{aligned} \left|\tilde{r}_{x,y}\right| &\ge r_{\min} \\ \left|\tilde{r}_{x,y}\right| &\ge r_2 - r_1 \end{aligned}\right\}, \qquad (1.3.16)$$

де найменше допустиме значення $r_{\min}$ для оцінки $\tilde{r}_{x,y}$ виражається формулою (1.8.12); $r_2 - r_1$ - довжина довірчого інтервалу $I_\beta = (r_1; r_2)$ для оцінки $\tilde{r}_{x,y}$ при довірчій імовірності $\beta \to 1$. Довірчий інтервал $I_\beta = (r_1; r_2)$ визначається методом Фішера, використовуючи спеціальну функцію виду

$$\tilde{Z} = \frac{1}{2} \ln \frac{1 + \left|\tilde{r}_{x,y}\right|}{1 - \left|\tilde{r}_{x,y}\right|}, \qquad (1.3.17)$$

яка підпорядковується нормальному законові розподілу. Для значень функції (1.3.17) складено таблицю (див. таблицю додатку 3). За значенням оцінки $\tilde{r}_{x,y}$ виражається величина $\tilde{Z}$ і визначається її довірчий інтервал $I_\beta(Z) = (Z_1; Z_2)$, де $Z_1 = \tilde{Z} - t_\beta \sigma_{\tilde{Z}}$, $Z_2 = \tilde{Z} + t_\beta \sigma_{\tilde{Z}}$. Тут $t_\beta = \arg\Phi^*\left(\frac{1+\beta}{2}\right)$ - параметр нормального закону розподілу, який визначається з таблиці нормальної функції розподілу (див. додаток 2 метод. вказівок **05-04-30**) для вибраної довірчої ймовірності β, а середнє квадратичне відхилення $\sigma_{\tilde{Z}}$ величини $\tilde{Z}$ обчислюється за формулою $\sigma_{\tilde{Z}} = \frac{1}{\sqrt{n-3}}$. Границі r_1 та r_2

довірчого інтервалу $I_\beta = (r_1; r_2)$ визначаються тепер за відповідними границями Z_1 та Z_2 довірчого інтервалу $I'_\beta = (Z_1; Z_2)$ із розв'язування оберненої задачі на основі функції (1.3.17) чи її таблиці.

Якщо лінійна імовірнісна залежність випадкових величин системи *(X,Y)* встановлена і це підтверджено умовами (1.3.8) чи (1.3.16), то таку залежність можна виразити аналітичною формою у вигляді лінійної емпіричної формули. Емпірична формула, яка виражає лінійний кореляційний зв'язок між величинами *X* та *У*, називається рівнянням регресії. Рівняння регресії мають вигляд:

$$\left.\begin{aligned} y_i - m_y &= \rho_{y/x}(x_i - m_x) \\ x_i - m_x &= \rho_{x/y}(y_i - m_y) \end{aligned}\right\} \qquad (1.3.18)$$

або відповідно:

$$\left.\begin{aligned} y_i &= \rho_{y/x} x_i + (m_y - \rho_{y/x} m_x) \\ x_i &= \rho_{x/y} y_i + (m_x - \rho_{x/y} m_y) \end{aligned}\right\}. \qquad (1.3.19)$$

Тут $\rho_{y/x} = r_{x,y}\dfrac{\sigma_y}{\sigma_x}$ - коефіцієнт регресії *У* на *X*; $\rho_{x/y} = r_{x,y}\dfrac{\sigma_x}{\sigma_y}$ - коефіцієнт регресії *X* на *У*. Середні квадратичні відхилення коефіцієнтів регресії обчислюють за формулами:

$$\sigma_{\rho_{x/y}} = \frac{\sigma_x}{\sigma_y}\sqrt{\frac{1-r_{x,y}^2}{n-3}}\,;\ \sigma_{\rho_{y/x}} = \frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-r_{x,y}^2}{n-3}}\,.$$

Рівняння регресії дозволяють розраховувати можливі значення величини *У*, якщо відомо значення величини *X*, і навпаки, виходячи з тісноти встановленої лінійної імовірнісної залежності між цими величинами. Ці рівняння апроксимують всі результати, які отримані в *n* проведених випробуваннях, і зображені на діаграмі залежності (малюнок 1.3.1). Для наочності встановлений в системі *(X,Y)* кореляційний зв'язок з тіснотою $r_{x,y}$ у формі лінійної регресії (1.3.19) можна показати графічно на одному малюнку разом з діаграмою залежності.

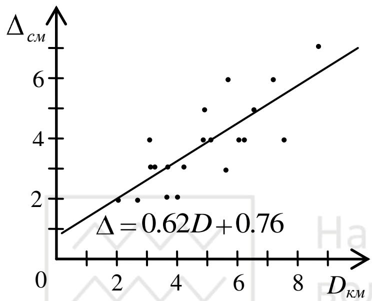
Порядок виконання роботи

Під час виконання роботи проводиться кореляційний аналіз результатів геодезичних вимірів у системі двох випадкових величин з метою встановлення залежності цих величин між собою і визначення тісноти та форми такої залежності.

Завдання 3. В результаті *n*=20 проведених польових вимірів довжин сторін мережі трилатерації отримано значення *D* виміряних довжин і відповідні кожній довжині помилки $|\Delta|$. З метою встановлення залежності помилок $|\Delta|$ від довжин *D* необхідно провести кореляційний аналіз цих величин. Вихідні дані за номером варіанту *N* студента вибираються з таблиці додатку 5.

Приклад. Нехай маємо наступні пари значень довжин D і помилок $|\Delta|$:

№	D (*км*)	$\lvert\Delta\rvert$ (*см*)	№	D (*км*)	$\lvert\Delta\rvert$ (*см*)	№	D (*км*)	$\lvert\Delta\rvert$ (*см*)	№	D (*км*)	$\lvert\Delta\rvert$ (*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



Малюнок 1.1.4

1. Будуємо діаграму залежності довжин D і помилок $|\Delta|$. Для цього вздовж осі абсцис позначуємо значення довжин, вздовж осі ординат – відповідні їм значення помилок, а на перетині ставимо точки (див. малюнок 1.1.4). Помилку $|\Delta|$ тут і далі для зручності позначуємо Δ.

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

2. Визначимо тісноту залежності значень довжин D і помилок Δ. Оскільки число вимірів $n < 50$, то тіснота залежності приблизно виражатиметься оцінкою коефіцієнта кореляції $\tilde{r}_{D,\Delta}$ і буде обчислюватись за формулою (1.3.11) з використанням формул (1.3.12)-(1.3.15) та (1.3.4)-(1.3.7). Результати проміжних розрахунків розмістимо в таблиці 2.

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

3. Визначення точності та надійності оцінки коефіцієнту кореляції $\tilde{r}_{D,\Delta}$ і встановлення кореляційного зв'язку між довжинами D і помилками Δ.

За таблицею додатку 3 для значень функції (1.3.17) при $\tilde{r}_{D,\Delta} = 0.78$ знаходимо $\tilde{Z} = 1.0454$. Середнє квадратичне відхилення $\sigma_{\tilde{Z}} = 0.243$. Значення параметру $t_\beta = \arg\Phi^*(\frac{1+\beta}{2})$ знайдемо з таблиці нормальної функції розподілу при довірчій імовірності $\beta = 0.9$: $t_\beta = \arg\Phi^*(0.95) = 1.645$.

Таблиця 2

Результати проміжних розрахунків

Результати вимірів			Результати розрахунків				
№	D_i	Δ_i	$D_i - \tilde{m}_D$	$\Delta_i - \tilde{m}_\Delta$	$(D_i - \tilde{m}_D)^2$	$(\Delta_i - \tilde{m}_\Delta)^2$	$(D_i - \tilde{m}_D)$x $(\Delta_i - \tilde{m}_\Delta)$
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

Отже, з ймовірністю 0.9 величина Z може набути значень $\tilde{Z} - t_\beta \sigma_{\tilde{Z}} \le Z \le \tilde{Z} + t_\beta \sigma_{\tilde{Z}}$, а її довірчий інтервал $I'_\beta =$ =(0.6457; 1.4451). За значеннями Z_1=0.6457 та Z_2=1.4451 шляхом оберненого табличного інтерполювання (див. таблицю додатку 3) знаходимо границі r_1 та r_2 довірчого інтервалу для коефіцієнту кореляції: $I_\beta = (r_1; r_2) =$ =(0.57; 0.89). Отже, з надійністю β= 0.9 точність невідомого значення коефіцієнту кореляції $r_{D,\Delta}$ окреслюється границями довірчого інтервалу I_β.

Довжина довірчого інтервалу $r_2 - r_1 =$ 0.32. Найменше допустиме значення $r_{\min}$ оцінки $\tilde{r}_{D,\Delta}$ при n =20 обчислюємо за формулою (1.3.9): $r_{\min}$ =0.50. Тепер перевіримо виконання вимог (1.3.16):

$$\left.\begin{array}{l} |\tilde{r}_{D,\Delta}| \ge r_{\min} \\ |\tilde{r}_{D,\Delta}| \ge r_2 - r_1 \end{array}\right\} \qquad \left.\begin{array}{l} 0.78 > 0.50 \\ 0.78 > 0.32 \end{array}\right\}.$$

Оскільки вимоги нерівностей (1.3.16) задовольняються, то прямолінійний кореляційний зв'язок між виміряними довжинами D та їх помилками Δ слід вважати встановленим.

4. Оскільки кореляційний зв'язок між довжинами та їх помилками існує, то виразимо його аналітичною формою у вигляді лінійного рівняння регресії (1.3.19). В поставленому завданні є зміст записувати лише рівняння регресії Δ на D виду $\Delta_i = \rho_{\Delta/D} D_i + (\tilde{m}_\Delta - \rho_{\Delta/D}\tilde{m}_D)$, де $\rho_{\Delta/D} = \tilde{r}_{D,\Delta}\dfrac{\tilde{\sigma}_\Delta}{\tilde{\sigma}_D} = =0.62$ - коефіцієнт регресії Δ на D. Остаточно $\Delta_i = 0.62 D_i + 0.76$.

Середнє квадратичне відхилення коефіцієнту регресії $\sigma_{\rho_{\Delta/D}} = \dfrac{\tilde{\sigma}_\Delta}{\tilde{\sigma}_D}\sqrt{\dfrac{1-\tilde{r}^2_{D,\Delta}}{n-3}} = 0.12$; $\rho_{\Delta/D} = 0.62 \pm 0.12$.

Рівняння регресії $\Delta_i = 0.62 D_i + 0.76$ зображено графічно у вигляді прямої на малюнку 1.3.2.

Отримане рівняння виражає лінійну імовірнісну залежність між заданими в завданні довжинами D та їх помилками Δ за відповідного комплексу умов. Воно дає можливість розраховувати ймовірні помилки вимірів ліній для довільних значень їх довжин при збереженні того комплексу умов, за яких отримані вихідні результати.

Рівняння регресії D на Δ формально може бути побудоване, однак задача визначення довжини лінії за помилкою її виміру не має практичного змісту.

Питання для самостійної роботи

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 коефіцієнт кореляції?
4. Коли і як обчислюється оцінка коефіцієнту кореляції?
5. Як визначити точність та надійність оцінки коефіцієнту кореляції?
6. Як можна зробити попередній висновок про існування кореляційного зв'язку між випадковими величинами?
7. Як виразити тісноту лінійної імовірнісної залежності в системі двох випадкових величин?
8. Як встановити існування кореляційного зв'язку між випадковими величинами при числі випробувань $n \geq 50$?
9. Як встановити існування кореляційного зв'язку між випадковими величинами при числі випробувань $n < 50$?
10. Що називають рівнянням регресії? Як записати таке рівняння?
11. Що виражає рівняння регресії і як воно використовується на практиці?
12. Як пояснити поняття „тіснота” та „форма” кореляційного зв'язку в системі двох випадкових величин?

РОЗПОДІЛ БАЛІВ,
що присвоюються студентам за виконання практичних робіт на тему:

№ з/п	Назви завдань	Кількість балів	
		Практичні завдання	Тести
1	Визначення закону розподілу на основі дослідних даних	2	2
2	Визначення точності та надійності числових оцінок параметрів розподілу при обмеженому числі випробувань	1	2
3	Визначення тісноти та форми кореляційного зв'язку в системі двох випадкових величин	1	2

ПЕРЕЛІК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Войтенко С.П. Математична обробка геодезичних вимірів. Теорія похибок вимірів: Навчальний посібник. – К.: КНУБА, 2003. – 216с.
2. Зазуляк П.М., Гавриш В.І., Євсєєва Е.М., Йосипчук М.Д. Основи математичного опрацювання геодезичних вимірювань: Підручник. – Львів: Растр-7, 2007. – 408 с.
3. Вентцель Е.С. Теория вероятностей. М., Наука, 1969. – 576 с.
4. Гайдаев П.А., Большаков В.Д. Теория математической обработки геодезических измерений. М., Недра, 1969. – 367с.
5. Видуев Н.Г., Григоренко А.Г. Математическая обработка геодезических измерений. Киев, Высшая школа, 1978.

ДОДАТКИ

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 параметру 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 3

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 4

Вихідні дані для виконання завдання №1

N	-4;-3	-3;-2	-2;-1	-1;0	0;1	1;2	2;3	3;4	N	-4;-3	-3;-2	-2;-1	-1;0	0;1	1;2	2;3	3;4	N	-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	51	9	35	78	129	140	112	48	14
2	10	49	78	145	130	75	49	12	27	8	20	31	48	59	45	25	9	52	13	28	80	131	127	72	30	9
3	9	44	73	120	138	88	33	9	28	10	15	30	57	50	39	21	17	53	17	24	75	111	120	104	45	14
4	10	29	71	140	128	79	34	12	29	8	20	34	54	62	48	19	10	54	16	19	56	112	103	71	32	11
5	11	44	89	125	143	82	34	8	30	8	18	29	54	51	40	31	12	55	17	20	51	92	101	81	39	13
6	14	51	88	151	140	73	40	11	31	10	24	54	83	89	38	27	15	56	8	22	61	94	76	51	26	9
7	15	29	84	134	145	81	35	17	32	16	35	74	129	136	84	38	14	57	9	26	34	71	82	61	38	10
8	10	44	77	119	129	80	27	16	33	14	32	90	138	129	82	42	13	58	12	24	55	73	62	49	27	8
9	11	25	48	88	92	34	26	12	34	17	36	72	138	128	83	40	12	59	11	32	48	59	66	42	35	8
10	9	22	43	76	69	36	24	8	35	23	90	131	145	129	84	39	10	60	12	26	47	61	51	38	20	15
11	10	24	48	85	71	33	25	9	36	12	31	86	125	119	83	40	15	61	8	18	31	49	60	52	38	12
12	10	25	52	91	94	35	24	12	37	15	48	87	150	132	78	51	14	62	8	19	58	69	50	39	18	10
13	9	21	54	96	91	30	20	12	38	10	27	79	124	131	80	32	9	63	12	19	32	64	78	55	39	11
14	9	23	56	90	84	28	19	10	39	11	41	69	130	118	68	25	13	64	9	20	63	88	74	39	28	13
15	17	44	91	130	142	80	38	20	40	16	52	92	161	140	81	49	15	65	11	28	58	92	96	72	51	10
16	16	39	95	164	172	87	43	7	41	12	45	77	120	134	80	33	10	66	13	30	82	109	96	61	38	11
17	12	31	89	150	140	86	41	16	42	10	23	49	88	80	39	25	11	67	8	30	61	101	118	88	49	9
18	9	42	82	121	138	92	38	10	43	9	14	78	135	146	80	35	15	68	7	31	88	126	109	68	40	8
19	8	34	85	132	119	76	30	11	44	8	23	59	98	89	36	28	11	69	10	26	71	128	132	91	29	9
20	10	29	75	115	108	66	40	15	45	16	40	88	129	142	106	78	22	70	11	42	86	145	131	78	30	7
21	9	40	61	96	107	59	31	8	46	14	70	132	182	169	91	59	11	71	14	40	91	134	158	120	41	10
22	11	32	68	101	96	71	42	16	47	13	64	120	166	178	120	54	12	72	13	40	102	169	140	91	62	11
23	19	36	61	86	93	52	28	15	48	12	58	122	160	151	95	69	16	73	10	32	72	149	166	112	40	12
24	12	29	54	84	77	59	30	18	49	11	38	69	135	154	109	50	13	74	8	29	120	181	163	108	55	13
25	9	25	49	69	76	48	21	10	50	10	34	89	142	130	96	58	15	75	10	20	89	164	183	170	96	14

Продовження таблиці додатку 4

N	-4;-3	-3;-2	-2;-1	-1;0	0;1	1;2	2;3	3;4	**N**	-4;-3	-3;-2	-2;-1	-1;0	0;1	1;2	2;3	3;4	**N**	-4;-3	-3;-2	-2;-1	-1;0	0;1	1;2	2;3	3;4
76	15	50	92	155	138	79	48	13	81	9	25	46	68	62	48	33	11	86	9	31	78	128	142	104	45	12
77	11	50	79	146	131	76	50	14	82	8	21	30	50	59	42	25	9	87	13	27	80	132	126	74	32	9
78	8	43	72	119	137	87	32	8	83	12	15	33	57	49	39	21	15	88	15	24	76	111	119	94	43	13
79	10	30	75	143	128	79	33	12	84	11	19	35	56	62	47	19	10	89	14	19	58	112	113	72	30	11
80	11	42	88	125	140	82	34	10	85	7	16	29	59	53	40	27	11	90	17	20	53	92	104	80	39	14

Додаток 5

Вихідні дані для виконання завдання №3

N	D(км)	$\|\Delta\|$ (см)	N	D(км)	$\|\Delta\|$ (см)	N	D(км)	$\|\Delta\|$ (см)	N	D(км)	$\|\Delta\|$ (см)	N	D(км)	$\|\Delta\|$ (см)	N	D(км)	$\|\Delta\|$ (см)
1	8.7	7.0	16	8.4	7.0	31	6.3	5.0	46	3.8	3.0	61	6.4	4.0	76	7.3	6.0
2	5.4	3.0	17	2.7	3.0	32	4.7	4.0	47	7.2	6.0	62	8.6	7.0	77	2.8	3.0
3	3.7	3.0	18	3.2	3.0	33	3.5	3.0	48	5.6	5.0	63	2.4	3.0	78	8.1	7.0
4	6.8	4.0	19	4.9	4.0	34	5.6	3.0	49	5.1	4.0	64	5.3	3.0	79	8.5	7.0
5	7.7	5.0	20	4.6	3.0	35	7.6	4.0	50	2.7	2.0	65	8.7	6.0	80	7.5	4.0
6	6.0	4.0	21	3.1	4.0	36	3.9	4.0	51	7.6	5.0	66	3.6	3.0	81	3.4	3.0
7	3.2	4.0	22	2.7	2.0	37	4.2	3.0	52	7.5	6.0	67	4.5	3.0	82	5.5	4.0
8	2.5	3.0	23	4.1	3.0	38	6.6	5.0	53	2.1	3.0	68	5.9	4.0	83	5.6	4.0
9	3.3	3.0	24	3.7	2.0	39	3.7	3.0	54	4.4	3.0	69	7.1	6.0	84	7.6	5.0
10	6.5	4.0	25	3.9	3.0	40	2.0	2.0	55	6.9	5.0	70	2.6	3.0	85	7.9	5.0
11	5.1	4.0	26	7.4	6.0	41	4.1	3.0	56	4.3	3.0	71	8.8	7.0	86	3.8	4.0
12	5.8	4.0	27	5.7	6.0	42	8.3	7.0	57	8.2	7.0	72	2.2	3.0	87	7.2	5.0
13	7.9	6.0	28	5.0	4.0	43	4.0	2.0	58	6.7	5.0	73	6.2	4.0	88	5.6	4.0
14	6.1	4.0	29	8.9	7.0	44	7.8	6.0	59	5.2	4.0	74	4.8	3.0	89	6.1	5.0
15	7.2	6.0	30	4.9	5.0	45	6.5	5.0	60	2.3	3.0	75	2.9	3.0	90	7.2	5.0