2. ГЕНЕРАЛЬНА ТА ВИБІРКОВА СУКУПНОСТІ, ЇХ ХАРАКТЕРИСТИКИ. ОЦІНКА ІСТИННОГО ЗНАЧЕННЯ ВИМІРЮВАНОЇ ВЕЛИЧИНИ. МАТЕМАТИЧНА ОБРОБКА РЕЗУЛЬТАТІВ ПРЯМИХ І НЕПРЯМИХ ВИМІРЮВАНЬ

2.1. Події, види подій, імовірність випадкових подій

Для вивчення реальних явищ чи процесів у мікро- чи макросвіті, в тому числі й у медико-біологічних системах, проводять дослідження. ***Дослідження*** – це науковий метод, який передбачає реалізацію комплексу певних умов. Результати досліджень невідомі заздалегідь, на математичній мові їх називають ***подіями***.

Події поділяють на:

1) ***достовірні*** – ті, які завжди відбуваються;
2) ***неможливі*** – ті, які ніколи не відбуваються;
3) ***випадкові*** – ті, які можуть відбутися, або не відбутися при даному дослідженні, напр. виграш у лотереї, випадання певного числа на гральному кубику.

Розділ математики, який називають «теорія ймовірностей», вивчає випадкові події, зокрема розраховує ймовірності їх появи. ***Ймовірністю події*** *А* називається відношення кількості сприятливих для даної події випадків *m* до числа випадків *n*, що можуть реалізуватися: $P(A)=\frac{m}{n}$. Це ***класичне визначення ймовірності***, яке безпосередньо не пов’язане з експериментальними дослідженнями, воно дає можливість вивчати об’єкт досліджень теоретично.

Ймовірність випадкової події є додатне число в межах від 0 до 1:

$$0 \le P(A) \le 1, \text{ оскільки } 0 \le m \le n.$$

Розглянемо серію з n досліджень, проведених в однакових умовах з метою вивчення події *А*. Якщо подія *А* реалізувалася m разів, то відношення числа дослідів, у яких подія *А* реалізувалася, до загального числа дослідів дасть нам частоту події: $\nu(A)=\frac{m}{n}$.

Статистичне визначення ймовірності: ймовірністю *P(A)* випадкової події *А* називають границю, до якої наближається частота події *А* при необмеженому зростанні повного числа випробувань: $P(A)=\lim\limits_{n\to\infty}\frac{m}{n}$.

Формула для статистичного визначення ймовірності використовується тоді, коли досліди були проведені фактично. Якщо кількість дослідів велика, то статистична імовірність практично дорівнює теоретично розрахованій ймовірності.

Розділ прикладної математики, в якому методами теорії ймовірностей досліджують результати спостережень та експериментів (статистичні дані) з метою одержання наукових і практичних висновків, називається ***математичною статистикою***. Методи статистичного аналізу використовуються при постановці експериментів для складання методично обґрунтованого плану дослідів, правильного їх проведення і отримання об'єктивних висновків.

2.2. Генеральна сукупність і вибірка

У біологічній і медичній статистиці досліджується велика кількість окремих об'єктів, що відрізняються один від одного і в той же час мають деякі подібні істотні ознаки (напр., кількість дітей, що народилася в місті за рік; кількість осіб, що мають певне захворювання та ін.).

Вся сукупність однорідних об'єктів, що досліджується відносно деякого якісного чи кількісного параметра, називається ***генеральною сукупністю***. Множина з n об'єктів, відібраних випадковим чином з генеральної сукупності, називається ***вибірковою сукупністю*** чи ***вибіркою*** (n – об'єм вибірки).

Дослідження статистичного матеріалу проводять за допомогою вибірки тому, що:

– дослідження всієї генеральної сукупності трудомістке, потребує великих затрат засобів і часу, чи практично нездійсненне;

– в окремих випадках дослідження всіх об'єктів вибірки привело б до їх псування (напр., контроль вмісту всіх ампул з лікарським препаратом в упаковці).

Метод, який ґрунтується на тому, що за даними обстеження вибірки, виділеної з генеральної сукупності, здійснюється висновок про всю генеральну сукупність, називається ***вибірковим методом***.

Вибірка називається ***репрезентативною***, якщо кожен об'єкт генеральної сукупності має однакову можливість попасти у вибірку.

Нехай із генеральної сукупності утворена вибірка, де x_1 спостерігалося n_1 разів, $x_2 = n_2$ разів, $x_k = n_k$ разів, і $n_1 + n_2 + ... + n_k = n$ – **об'єм вибірки**. Значення $x_1, x_2, ... x_k$, які спостерігаються, називаються ***варіантами***. Послідовність варіант у порядку зростання називається ***варіаційним рядом***. $n_1, n_2, n_3, ... n_k$ – ***частоти***, кількість випадків, у яких спостерігалися варіанти.

2.3. Дискретні та неперервні випадкові величини

Дискретною випадковою величиною називається величина, яка може приймати лише певний ряд значень і не може приймати проміжних.

Цей ряд значень може бути обмеженим або необмеженим (наприклад: натуральні числа). Кожному значенню можна приписати певний номер і їх можна перелічити. Тому множина значень дискретної величини є зліченна.

Для випадку, коли кількісна ознака є дискретною (напр., кількість хворих у лікарні, кількість таблеток в упаковці), тобто може приймати тільки певні значення, підраховують, скільки разів зустрічається кожне значення і результат представляють у вигляді таблиці:

X	x_1	x_2	...	x_i	...	x_k
n	n_1	n_2	...	n_i	...	n_k

Перший рядок містить всі різні значення варіант X у порядку їх зростання чи спадання, другий – числа n, що дорівнюють числу варіант із значенням X. Така таблиця називається *статистичним дискретним рядом розподілу.*

> ***Неперервна випадкова величина*** – це величина, яка на будь-якому малому інтервалі може приймати безліч значень.

Ці значення пронумерувати і перелічити неможливо, тому множину значень неперервної величини називають незліченною.

Якщо кількісна ознака неперервна (напр. значення артеріального тиску, температура тіла) і наявна велика кількість варіант, статистичний розподіл задають у вигляді послідовності інтервалів і відповідних їм частот:

Інтервал x	$\left]x_0, x_i\right[$	$\left]x_1, x_2\right[$	...	$\left]x_{k-1}, x_k\right[$
n	n_1	n_2	...	n_k

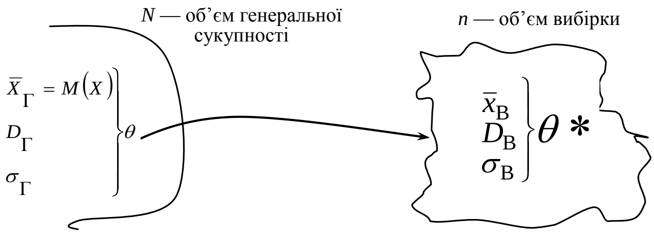
Така таблиця називається *інтервальним рядом розподілу*.

2.4. Оцінка параметрів генеральної сукупності за її вибіркою. Точкове оцінювання

> Узагальнення властивостей ознак вибірки стосовно генеральної сукупності називають ***статистичним висновком***.

Інформація про ознаку генеральної сукупності, яку отримали на основі обробки вибірки, завжди міститиме певні похибки, оскільки вибірка становить лише незначну частину від неї ($n << N$), тобто об'єм вибірки значно менший від об'єму генеральної сукупності.

Тому вибірку утворюють так, щоб отримана інформація була найповнішою (вибірка має бути репрезентативною) і забезпечувала з найбільшим ступенем довіри дані про параметри генеральної сукупності. Параметри генеральної сукупності, а саме: генеральне середнє $\overline{X}_{Г}$, генеральна дисперсія $D_{Г}$, генеральне середнє квадратичне відхилення $\sigma_{Г}$ та ін. є величинами сталими, але їх числові значення невідомі. Ці параметри оцінюються параметрами вибірки: вибіркове середнє $\bar{x}_{B}$, вибіркова дисперсія D_{B}, вибіркове середнє квадратичне відхилення σ_{B}, які отримують при дослідженні вибірки. Параметри вибірки є величинами непередбачуваними, тобто випадковими. Схематично це можна показати так:



Тут через θ позначено оцінювальний параметр генеральної сукупності, а через $\theta*$ — його статистичну оцінку, яку називають ще *статистикою*. При цьому $\theta = const$, а $\theta*$ — випадкова величина, що має певний закон розподілу ймовірностей.

Оцінюванням у статистиці називають встановлення значення параметра розподілу (або функції від нього) за даними спостережень (одержання характеристик розподілу генеральної сукупності на основі даних вибірки).

Числові характеристики вибірок є оцінками відповідних числових характеристик генеральних сукупностей випадкових величин.

Оцінка параметрів генеральної сукупності за характеристиками вибірки – одна з основних проблем, яка виникає при статистичному аналізі даних.

Оцінювання може бути *точковим* та *інтервальним*. *Точкове оцінювання* дає орієнтовне числове значення характеристики генеральної сукупності. *Інтервальне оцінювання* вказує на проміжок, у якому з наперед заданою надійністю має перебувати характеристика генеральної сукупності.

Точкове оцінювання

При точковому оцінюванні отримуємо орієнтовне числове значення характеристики розподілу генеральної сукупності (параметра генеральної сукупності), яка обчислюється для вибірки.

Статистика для визначення параметра розподілу генеральної сукупності має назву ***точкової оцінки даного параметра***.

Величина відхилення характеристики вибірки (статистики) від відповідного значення характеристики генеральної сукупності називається ***статистичною похибкою*** (***стандартною похибкою***) або ***похибкою репрезентативності***.

Статистичні похибки належать тільки характеристикам вибірки, вони виникають у процесі відбору варіант із генеральної сукупності.

Статистичні похибки характеризують варіювання показників вибірки навколо відповідних характеристик генеральної сукупності, тобто вказують на точність, з якою показник вибірки репрезентує параметр генеральної сукупності. Вони володіють тими ж властивостями, що і середнє квадратичне відхилення. Чим сильніше варіює ознака, тим більша при рівних умовах буде похибка показників вибірки, і, навпаки, при слабому варіюванні ознаки похибка показників вибірки буде меншою.

Специфічна властивість статистичних похибок

Статистичні похибки зменшуються при збільшенні об’єму вибірки. Ця властивість статистичних похибок обумовлена дією закону великих чисел, згідно з яким найбільш ймовірнісний результат отримується при найбільшій кількості спостережень. Чим менша похибка, тим ближче величина характеристики вибірки наближається до величини параметра генеральної сукупності, і, навпаки, чим більша похибка, тим менш точно характеристика вибірки репрезентує параметр генеральної сукупності.

Стандартні похибки основних статистик (оцінок):

Похибка середнього арифметичного: $s_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ (2.1)

Похибка дисперсії: $s_{\sigma^2} = \dfrac{\sigma^2}{\sqrt{2n}}$ (2.2)

Похибка середнього квадратичного відхилення: $s_{\sigma} = \dfrac{\sigma}{\sqrt{2n}}$ (2.3)

де n – об’єм вибірки, σ - середнє квадратичне відхилення вибірки.

До точкових оцінок висуваються такі вимоги:

1. **Незміщеність (точність)**. Означає, що при проведенні великої кількості спостережень із вибірками однакового розміру середнє значення кожної вибірки прямує до істинного значення генеральної сукупності. Зміщеність обумовлюється наявністю *статистичної похибки* або *похибки репрезентативності.*

Математичне сподівання незміщеної оцінки співпадає із значенням параметра генеральної сукупності. Математичне сподівання дискретної випадкової величини визначається як сума добутків значень випадкової величини на ймовірності появи цих значень.

2. **Спроможність.** Із зростанням об'єму вибірки оцінка повинна прямувати до значення відповідного параметра генеральної сукупності з ймовірністю, яка прямує до 1.

3. **Ефективність.** Вибрана оцінка повинна мати мінімальну дисперсію у порівнянні з іншими аналогічними оцінками, тобто повинна виявляти найменшу випадкову варіацію.

4. **Достатність.** Оцінка повинна містити всю необхідну інформацію і не вимагати додаткової.

Оцінка, що бездоганна за деяких припущень щодо вихідних даних, при відхиленні від цих припущень може спричинити значне викривлення результатів. Наприклад, вибіркове середнє має багато властивостей оптимальності для нормально розподілених вибірок, але стає неспроможним за умови наявності у вибірці промахів, тобто значень, що різко відрізняються від інших. Зазвичай такі промахи зумовлені похибками вимірювань, які за допомогою простих прийомів попереднього аналізу даних вилучають із вибірки.

Вибіркова середня є незміщеною і спроможною оцінкою генеральної середньої.

$$\bar{x}_B = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{2.4}$$

Вибіркова дисперсія, яка розрахована за формулою:

$$D_B = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_B)^2 \tag{2.5}$$

є зміщеною по відношенню до свого генерального параметра на величину *n/(n-1)*. Щоб отримати незміщену дисперсію, потрібно у формулу ввести як множник поправку на зміщеність.

«Виправлена дисперсія» S^2 обчислюється за формулою:

$$S^2 = \frac{n}{n-1}D_B = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_B)^2. \tag{2.6}$$

Тому для оцінки генеральної дисперсії приймають виправлену дисперсію.

Для оцінки **середнього квадратичного відхилення** генеральної сукупності використовують виправлене середнє квадратичне відхилення, яке дорівнює квадратному кореню з виправленої дисперсії.

$$S=\sqrt{S^2}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}_B)^2} \qquad (2.7)$$

Очевидно, що при великих значеннях об'єму вибірки n виправлена і вибіркова дисперсії відрізняються мало. На практиці користуються виправленою дисперсією, якщо n<30.

Приклад. При дослідженні сироватки крові на холестерин (мг %) у чоловіків (46-50 років) під час гіпертонічного кризу одержані такі результати: 210, 215, 230, 231, 232, 231, 238, 240, 245. Обчислити вибіркове середнє, дисперсію, стандартне відхилення для середнього арифметичного.

Розв'язання. Вибіркове середнє обчислюємо за формулою:

$$\bar{x}_в=\frac{\sum_{i=1}^{к} m_i x_i}{n}=\frac{1}{9}(210+215+230+2\cdot 231+232+238+240+245)=230 \ (мг\ \%).$$

Знайдемо виправлену дисперсію:

$$S^2=\frac{\sum_{i=1}^{k} m_i(x_i-\bar{x}_в)^2}{n-1}=\frac{1}{9-1}\cdot((210-230)^2+(215-230)^2+(230-230)^2+$$
$$+2\cdot(231-230)^2+(232-230)^2+(238-230)^2+(240-230)^2+(245-230)^2)=$$
$$=127{,}44(мг\%)^2.$$

Виправлене середнє квадратичне відхилення обчислимо за формулою

$$S=\sqrt{S^2}=11{,}29(мг\%)\ .$$

Середнє квадратичне відхилення вибіркового середнього отримаємо за формулою $S_{\bar{x}}=S/\sqrt{n}=11{,}29/\sqrt{9}=3{,}76\ (мг\%)$.

2.5. Інтервальне оцінювання

При невеликому об'ємі вибірки (n<30) користуються інтервальними оцінками, оскільки точкова оцінка, яка визначається одним числом, може призводити до грубих помилок.

Інтервальними називають оцінки, які визначаються двома числами – кінцями інтервалу, що покриває оцінюваний параметр. Інтервальні оцінки дозволяють встановити точність та надійність оцінок.

Інтервальне оцінювання вказує на проміжок, у якому з наперед заданою ймовірністю має перебувати характеристика розподілу генеральної сукупності (параметр генеральної сукупності).

Нехай за даними вибірки статистика θ^* служить оцінкою невідомого параметра генеральної сукупності θ (математичне очікування, дисперсія, стандартне відхилення). Очевидно, що θ^* тим точніше визначає параметр θ, чим менша абсолютна величина різниці $|\theta-\theta^*|$. Іншими словами, якщо $\delta>0$ і $|\theta-\theta^*|<\delta$, то чим менше δ, тим оцінка точніша. Отже, додатне число δ характеризує **точність оцінки**.

Різниця між статистичною оцінкою θ^* та її оцінювальним параметром θ, взята за абсолютним значенням, називається ***точністю оцінки***

$$|\theta-\theta^*|<\delta \qquad (2.8)$$

Статистичні методи не дозволяють категорично стверджувати, що оцінка θ^* задовольняє нерівність $|\theta-\theta^*|<\delta$, можна говорити тільки про ймовірність, з якою нерівність задовольняється.

Ймовірності, які визнані достатніми для впевненого твердження про параметри генеральної сукупності на базі відомих показників вибірки називають ***довірчими.***

Це поняття запропоноване Р.Фішером. *Воно витікає з принципу, який покладений в основу застосування теорії ймовірностей до вирішення практичних завдань*. Відповідно до цього, малоймовірні події вважаються практично неможливими, а події, ймовірність яких близька до одиниці, приймають за майже достовірні.

Оскільки θ^* є випадковою величиною, то і δ буде випадковою, тому нерівність (2.9) справджуватиметься з певною ймовірністю.

Довірчою ймовірністю (надійністю оцінки) називається ймовірність попадання параметра генеральної сукупності θ, який оцінюється, в інтервал $(\theta^* - \delta, \theta^* + \delta)$.

$$P(\theta^* - \delta < \theta < \theta^* + \delta) = \gamma$$

$$P(|\theta-\theta^*|<\delta)= \gamma \qquad (2.9)$$

де θ^* - величина статистики вибірки (величина випадкова);

θ - параметр, що оцінюється (величина постійна).

Інтервал $(\theta^* - \delta, \theta^* + \delta)$, який покриває з ймовірністю γ невідомий параметр, що оцінюється, називають ***довірчим інтервалом***.

Як правило, як довірчі використовують ймовірності P_1=0,95; P_2=0,99 і P_3=0,999. Це означає, що при оцінці генеральних параметрів за відомими показникам вибірки існує ризик помилитися в першому випадку один раз на 20 дослідів, у другому – один раз на 100 дослідів і в третьому – один раз на 1000 дослідів.

Наприклад, є дві вибірки однакового об’єму: одужання в п’ятиденний термін при певному захворюванні та відповідність стандартам продукції фармацевтичної фабрики лікарських засобів. Довірча ймовірність γ=0,99 стосовно першої вибірки означає, що із 100 пацієнтів лише один не одужає в п’ятиденний термін. Це непоганий показник. А для фармацевтичної фабрики – із 100 препаратів один бракований – зовсім погано, як γ, в цьому випадку, потрібно взяти більше число, наприклад, γ=0,999.

Рівень значущості

Якщо надійність оцінки γ=0,95, тоді число $\alpha=1-\gamma=0,05$ показує з якою ймовірністю висновок про надійність оцінки помилковий. Число α називається ***рівнем значущості***.

Рівень значущості – це та ймовірність, якою вирішено нехтувати в даному дослідженні.

Визначення довірчого інтервалу для середнього значення генеральної сукупності

Визначимо довірчий інтервал для оцінки середнього значення μ випадкової величини при невідомому середньому квадратичному відхиленні σ. У цьому випадку за даними вибірки будують розподіл випадкової величини *T* (можливі значення якої позначають *t*), який називається *t*-розподілом або ***розподілом Стьюдента*** (псевдонім англійського статистика У. Госсета).

$$T=\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}. \tag{2.10}$$

Генеральна середня μ з рівнем значущості α знаходиться в довірчому інтервалі $(\bar{x}-\delta;\bar{x}+\delta)$, де $\delta=t(\alpha,k)\cdot\frac{S}{\sqrt{n}}$, або

$$\bar{x}-t(\alpha,f)\cdot\frac{S}{\sqrt{n}}\le\mu\le t(\alpha,f)\cdot\frac{s}{\sqrt{n}} \tag{2.11}$$

де $\bar{x}$ та S - вибіркові (точкові) оцінки середнього і середнього квадратичного відхилень (знаходять за даними вибірки), $\delta=t\frac{s}{\sqrt{n}}$ - точність оцінки, *n* – розмір вибірки, $t(\alpha,f)$ – табличні значення коефіцієнта Стьюдента (див. додаток). Тут α - рівень значущості, *f=n-1* число ступенів вільності. Під числом ступенів вільності розуміють число змінних, значення яких задаються довільно. Фактично це є загальне число змінних мінус число лінійних зв'язків, накладених на систему, що вивчається.

Приклади

1. *При визначенні маси n=20 таблеток кордарону отримали, що вибіркова середня* $\bar{x} = 200$ *мг, “виправлене” середнє квадратичне відхилення S = 4,0 мг.*
Відомо, що маса таблетки має нормальний закон розподілу. Знайти за даними вибірки довірчий інтервал для середнього арифметичного з надійністю γ*=0,99.*

Розв’язання.

Для надійності γ=0,99 та n=20 знаходимо з таблиці $t_\gamma = 2{,}861$. Тоді

$$\delta = 2{,}861\frac{4{,}0}{\sqrt{20}} \approx 2{,}6 \cdot$$

Отже, довірчий інтервал (197,4; 202,6) включає з надійністю 0,99 середнє арифметичне генеральної сукупності.

2. *Маса насінини ріжкового дерева має нормальний закон розподілу. Знайти за даними вибірки довірчий інтервал для математичного сподівання* μ *з надійністю* γ*=0,99, якщо зважувались n=20 насінин, вибіркове середнє* $\bar{x} = 200$ *мг, «виправлене» середнє квадратичне відхилення S = 4,0 мг.*

Розв’язання.

Для надійності γ=0,99 та n=20 знаходимо за таблицею розподілу Стьюдента t_γ=2,861. Тому $\delta = 2{,}861 \cdot \frac{4{,}0}{\sqrt{20}} \approx 2{,}6$. Кінці довірчого інтервалу 200-2,6 та 200+2,6. Тоді довірчий інтервал для μ з надійністю 0,99: $197{,}4 < \mu < 202{,}6$.

2.6. Обробка результатів прямих вимірювань

Нехай проведено n прямих вимірів фізичної величини X, істинне значення якої позначимо μ. Покази приладу – $x_1, x_2,...,x_n$ – це числові значення X (напр. визначення температури за шкалою термометра чи визначення сили струму за показами амперметра).

Вважаємо, що всі вимірювання ***рівноточні***, тобто виконані одним методом в однакових умовах (їх дисперсії однакові).

Оцінками істинного значення прямо виміряних на досліді величин є:

1. ***Середнє вибіркове значення вимірюваної величини*** (середнє арифметичне з n - вимірювань)

$$\bar{x}_B = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \ . \qquad (2.12)$$

2. ***Виправлена вибіркова дисперсія*** вимірюваної величини (для числа вимірів n<30)

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_B)^2 \ . \qquad (2.13)$$

3. ***Виправлене середньоквадратичне відхилення***, що дорівнює середньоквадратичній похибці окремого вимірювання (стандарт)

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_B)^2} \ . \qquad (2.14)$$

4. Середня квадратична похибка середнього арифметичного

$$s_{\bar{x}_B} = \frac{s}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \bar{x}_B)^2} \ . \qquad (2.15)$$

5. ***Найбільш можлива похибка окремого вимірювання***:

$$s_{\max} = 3s \ . \qquad (2.16)$$

6. ***Точність прямого вимірювання*** (або абсолютна похибка середнього арифметичного незалежних вимірювань) – це абсолютна величина різниці між істинним значенням вимірюваної величини і середнім вибірковим

$$\Delta x = \left|\mu - \bar{x}_B\right| = t_\gamma(f)\frac{s}{\sqrt{n}} = t_\gamma(f)s_{\bar{x}_B} \ , (f=n\text{-}1), \qquad (2.17)$$

де $t_\gamma(f)$ - коефіцієнт Стьюдента, числові значення якого наведені в додатку .

7. Довірчий інтервал вимірюваної величини

$$\bar{x}_B - \Delta x < \mu < \bar{x}_B + \Delta x \qquad (2.18)$$

$$\bar{x}_B - t_\gamma(f)s_{\bar{x}_B} < \mu < \bar{x}_B + t_\gamma(f)s_{\bar{x}_B} \qquad (2.19)$$

Обробку серії прямих вимірювань проводять у такому порядку:

1) визначають середнє арифметичне (вибіркове);
2) знаходять середню квадратичну похибку окремого вимірювання;
3) визначають найбільшу можливу похибку s_{max} і переконуються, що серед результатів вимірювання нема таких, які відрізнялись би від середнього арифметичного більше, ніж на s_{max}. Якщо такі результати є, їх відкидають і обробку починають спочатку;
4) визначають середню квадратичну похибку середнього арифметичного;
5) визначають точність вимірювань Δx при заданій довірчій імовірності γ і кількості вимірювань n;
6) записують довірчий інтервал для вимірюваної величини.

Приклад. *Сім вимірювань діаметра капіляра в стінці легеневих альвеол дали такі результати: (у мікронах) 2,83; 2,82; 2,81; 2,82; 2,86; 2,83; 2,84. Провести обробку результатів цих вимірювань*.

Розв'язання.

1) $\overline{x}_B = \dfrac{2,83+2,82+2,81+2,82+2,86+2,83+2,84}{7} = 2,83мк$

2)

$$s = \sqrt{\frac{(2,83-2,83)^2+(2,82-2,83)^2+(2,81-2,83)^2+(2,82-2,83)^2+(2,86-2,83)^2+(2,83-2,83)^{22}+(2,84-2,83)^2}{6}} =$$

= 0,016 мк.

3) s_{max}=3·0,016=0,048 мк.

Результати всіх вимірювань не відрізняються від $\overline{x}_B$ більше ніж на 0,048 мк, тому нема потреби відкидати окремі результати.

4) $s_{\overline{x}_B} = \dfrac{s}{\sqrt{n}} = \dfrac{0,016}{\sqrt{7}} = 0,006$ мк

5) $\Delta x = t_\gamma(f)s_{\bar{x}_B} = 2,447 \cdot 0,006 = 0,014$ мк, оскільки у випадку γ=0,95 і *n*=7 $t_\gamma(f)$ =2,447.

Значення діаметра альвеол із врахуванням тільки випадкової похибки: x=(2,83 ± 0,014)мк.

6) довірчий інтервал має межі від 2,816 до 2,844. Тобто, з імовірністю γ=0,95, істинне значення діаметра альвеол не вийде за межі інтервалу $]2,816\ i\ 2,844[$.

2.7. Обробка результатів непрямих вимірювань

Розглянемо випадок, коли шукана величина y зв'язана з вимірюваною величиною x (або декількома вимірюваними величинами x_i) відомою функціональною залежністю:

$$y=f(x_1, x_2, \ldots, x_i, \ldots, x_k). \tag{2.21}$$

Такі вимірювання називаються непрямими.

Оцінкою істинного значення непрямо вимірюваної величини y є середнє значення:

$$\overline{y} = f(\overline{x_1}, \overline{x_2}, \ldots, \overline{x_k}), \tag{2.22}$$

де $\overline{x_1}, \overline{x_2}, \ldots, \overline{x_k}$ - середні арифметичні величин $x_1, x_2, \ldots, x_k$, що вимірюються.

Оцінкою середнього квадратичного відхилення є середня квадратична похибка середнього $\overline{y}$, що обчислюється за формулою:

$$s_{\bar{y}} = \sqrt{\left(\frac{\partial f}{\partial x_1} s_{\bar{x}_1}\right)^2 + \left(\frac{\partial f}{\partial x_2} s_{\bar{x}_2}\right)^2 + \ldots + \left(\frac{\partial f}{\partial x_k} s_{\bar{x}_k}\right)^2} \,. \qquad (2.23)$$

У цій формулі похідні $\frac{\partial f}{\partial x_i}$ (i = 1, 2, ..., k) повинні бути взяті в точках $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$, а $s_{\bar{x}_i}$ - це середні квадратичні похибки середніх значень вимірюваних величин. Їх обчислюють за формулою

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{k} m_i (x_i - \bar{x})^2} \qquad (2.24)$$

Формула (2.21) застосовується, якщо всі x_i мають незалежні випадкові похибки.

Абсолютна похибка $\Delta \bar{y}$ середнього значення $\bar{y}$ оцінюється за формулою:

$$\Delta \bar{y} = \left| \bar{y} - y \right| = t_\gamma (f) s_{\bar{y}} \,. \qquad (2.25)$$

Інтервальною оцінкою y є довірчий інтервал $\left| \bar{y} - \Delta \bar{y},\ \bar{y} + \Delta \bar{y} \right|$, в який попадає істинне значення непрямо виміряної величини y із заданою довірчою імовірністю γ. Кінцевий результат записують у вигляді:

$$y = \bar{y} \pm \Delta \bar{y} \,. \qquad (2.26)$$

Відносна похибка середнього значення $\bar{y}$:

$$\delta_{\bar{y}} = \frac{\Delta \bar{y}}{\bar{y}} \cdot 100\% \qquad (2.27)$$

Приклад. *З порошку амідопірину на гідропресі спресували таблетки. Зважили п'ять таблеток і виміряли їх товщину та діаметр. Результати вимірювань подали в таблиці:*

m_i, г	*0,388*	*0,390*	*0,387*	*0,389*	*0,388*
h_i, см	*0,53*	*0,54*	*0,53*	*0,55*	*0,54*
d_i, см	*0,92*	*0,92*	*0,93*	*0,93*	*0,91*

Визначити густину амідопірину в таблетках за формулою $\rho = \frac{4m}{\pi d^2 h}$, *де m – маса таблетки, h – її товщина, d – діаметр. Знайти оцінку істинної густини амідопірину, абсолютну і відносну похибки при довірчій імовірності 0,95.*

Розв'язання.

Середнє значення безпосередньо виміряних величин:

$$\overline{m} = \frac{0{,}388 + 0{,}390 + 0{,}387 + 0{,}389 + 0{,}388}{5} = 0{,}388 \text{ г};$$

$$\overline{h} = \frac{0{,}53 + 0{,}54 + 0{,}53 + 0{,}55 + 0{,}54}{5} = 0{,}54 \text{ см};$$

$$\overline{d} = \frac{0{,}92 + 0{,}92 + 0{,}93 + 0{,}93 + 0{,}91}{5} = 0{,}92 \text{ см}.$$

Середнє значення густини амідопірину

$$\overline{\rho} = \frac{4m}{\pi d^2 h} = \frac{4 \cdot 0{,}388}{3{,}14 \cdot 0{,}92^2 \cdot 0{,}54} = 1{,}09 \text{ г/см}^3.$$

За формулою (2.15) знаходимо оцінку середнього квадратичного відхилення середніх значень безпосередньо виміряних величин:

$$s_{\overline{m}} = \sqrt{\frac{0{,}002^2 + 0{,}001^2 + 0{,}001^2}{20}} = 5{,}5 \cdot 10^{-4} \text{ г};$$

$$s_{\overline{h}} = \sqrt{\frac{0{,}01^2 + 0{,}01^2 + 0{,}01^2}{20}} = 3{,}9 \cdot 10^{-3} \text{ см};$$

$$s_{\overline{d}} = \sqrt{\frac{0{,}01^2 + 0{,}01^2 + 0{,}01^2}{20}} = 3{,}9 \cdot 10^{-3} \text{ см}.$$

Згідно з формулою (2.21), оцінка середнього квадратичного відхилення середнього $\overline{\rho}$

$$s_{\overline{\rho}} = \sqrt{\left(\frac{\partial \rho}{\partial m} s_{\overline{m}}\right)^2 + \left(\frac{\partial \rho}{\partial h} s_{\overline{h}}\right)^2 + \ldots + \left(\frac{\partial \rho}{\partial d} s_{\overline{d}}\right)^2} =$$

$$= \sqrt{\left(\frac{4}{\pi d^2 h}\right)^2 s_{\overline{m}}^2 + \left(-\frac{4m}{\pi d^2 h^2}\right)^2 s_{\overline{h}}^2 + \left(\frac{8m}{\pi d^3 h}\right)^2 s_{\overline{d}}^2} = 1{,}218 \cdot 10^{-2} \, г/см^3$$

З додатку знаходимо значення коефіцієнта Стьюдента $t_{0,95}(4)=2{,}776$.

$$\Delta \overline{\rho} = 2{,}776 \cdot 12{,}18 \cdot 10^{-3} = 0{,}03 \; г/см^3$$

Відносна похибка

$$\delta = \frac{0{,}03}{1{,}09} \cdot 100\% = 2{,}7\% \, .$$

Істинне значення густини амідопірину $\rho = (1{,}09 \pm 0{,}03) г/см^3$.

Контрольні запитання

1. Події, види подій, ймовірності випадкових подій.
2. Поняття генеральної та вибіркової сукупностей.
3. Дискретні та неперервні випадкові величини.

4. Поняття статистичної оцінки. Довірча ймовірність.
5. Генеральне та вибіркове середнє.
6. Генеральна та вибіркова дисперсії.
7. Виправлена дисперсія та виправлене середнє квадратичне.
8. Оцінка середнього квадратичного відхилення вибіркової середньої.
9. Довірчий інтервал для оцінки середнього значення випадкової величини.
10. Математична обробка результатів прямих вимірювань.
11. Математична обробка результатів непрямих вимірювань

Завдання для самостійної практичної роботи:

1. З генеральної сукупності взято вибірку об'єму n=50:

Варіанти, x_i	2	5	7	10
Частоти, n_i	16	12	8	14

Знайти незміщену оцінку генерального середнього.

2. З генеральної сукупності взято вибірку об'єму n=60:

Варіанти, x_i	1	3	6	26
Частоти n_i	8	40	10	2

Знайти незміщену оцінку генерального середнього.

3. За вибіркою n=41 знайдена зміщена оцінка D_B=3 генеральної дисперсії. Знайти незміщену оцінку дисперсії генеральної сукупності.

4. За вибіркою n=51 знайдена зміщена оцінка D_B=5 генеральної дисперсії. Знайти незміщену оцінку дисперсії генеральної сукупності.

5. У результаті п'яти вимірювань довжини стержня одним приладом отримані такі результати (в мм): 92, 94, 103, 105, 106. Знайти вибіркову середню довжину стержня, вибіркову і виправлену дисперсію похибок приладу.

6. У результаті чотирьох вимірювань деякої фізичної величини отримані результати: 9, 8, 11, 12. Знайти вибіркову середню результатів вимірювання, вибіркову і виправлену дисперсію похибок приладу.

7. При визначенні маси 5 таблеток кетотифену були отримані такі результати: 0,00097; 0,00099; 0,0010; 0,0011; 0,0013. Знайти середню масу таблетки і середнє квадратичне відхилення вибіркової середньої.

8. За вибіркою були отримані дані про масу новонароджених морських свинок (в г): 30, 30, 25, 32, 30, 25, 33, 32, 29, 28, 27, 36, 31, 34, 30, 29, 23, 36, 30. Знайти вибіркову середню, вибіркову та виправлену дисперсію за даними вимірювання та середнє квадратичне відхилення вибіркової середньої

9. Дані про кількість квіток лікарської рослини на одному пагоні утворюють такий варіаційний ряд:

Кількість квіток	0	2	3	4	5	6	7	8
Частота	2	3	2	6	5	5	1	1

Обчислити вибіркове середнє, дисперсію, стандартне відхилення, стандартне відхилення для середнього.

10. Деяка випадкова величина має нормальний розподіл. Для вибірки n=20 вибіркове середнє $\bar{x}_в$ =15,6, а оцінка середньоквадратичного відхилення s=0,6. Визначити інтервальну оцінку математичного сподівання з довірчою ймовірністю γ=0,99.

11. Знайти, користуючись розподілом Стьюдента, довірчий інтервал для оцінки генеральної середньої μ із надійністю γ=0,99, якщо виправлене середнє квадратичне відхилення s=0,6, вибіркова середня $\bar{x}_в$ =20,2. Об'єм вибірки n=16.

12. Знайти, користуючись розподілом Стьюдента, довірчий інтервал для оцінки генеральної середньої μ із надійністю γ = 0,99, якщо виправлене середнє квадратичне відхилення s=2,4, вибіркова середня $\bar{x}_в$ =14,2. Об'єм вибірки n=9.

13. Знайти, користуючись розподілом Стьюдента, довірчий інтервал для оцінки генеральної середньої μ із надійністю γ = 0,95, якщо виправлене середнє квадратичне відхилення s=1,6, вибіркова середня $\bar{x}_в$ =16,8. Об'єм вибірки n=12.

14. Знайти, користуючись розподілом Стьюдента, довірчий інтервал для оцінки генеральної середньої μ із надійністю γ=0,95, якщо виправлене середнє квадратичне відхилення s=6,6, вибіркова середня $\bar{x}_в$ =15. Об'єм вибірки n=14.

15. Зроблено 12 вимірювань одним приладом, при цьому виправлене середнє квадратичне відхилення виявилось рівним 8. Знайти точність приладу з надійністю γ=0,95.

16. У результаті п'яти зважувань таблеток еритроміцину отримані такі дані (в мг): 96, 99, 100, 101, 104. Знайти вибіркову середню, дисперсію і виправлену дисперсію за результатами вимірювань.

17. Проводилися обстеження групи спортсменів із 15 чоловік. При вимірюванні обхвату грудної клітки встановлено, що у двох вона дорівнює 90 см, у трьох – 96, у двох – 97, у трьох – 98, у чотирьох –100, у одного – 102. Знайти вибіркову середню та середнє квадратичне відхилення вибіркової середньої.

18. Знайти вибіркову середню та дисперсію за даними вимірювання довжини хоботка (у мм) у 6 бджіл: 6,54; 6,71; 6,70; 6,69; 6,70; 6,62.

19. При перерахунку листя у однієї з лікарських рослин були отримані такі дані: 8, 10, 7, 9, 11, 6, 9, 8, 7, 7. Обчислити вибіркову середню і середнє квадратичне відхилення вибіркової середньої.

20. Вміст сульфату хініну в таблетках визначався спектрофотометричним

методом. У результаті шести вимірювань одержали такі результати: 99,9; 99,8; 99,6; 99,1; 99,2; 99,2 %. Визначити середнє значення при довірчій ймовірності γ=0,95.

21. У результаті десяти однакових проб були отримані такі значення вмісту калію (у %): 1; 1,05; 1,1; 0,99; 0,97; 0,98; 1,08; 1,07; 1,01; 1,03. Обчислити оцінку істинного вмісту калію при довірчій ймовірності γ=0,95.

22. У результаті десяти однакових проб були отримані такі значення вмісту марганцю (у %): 0,69; 0,70; 0,67; 0,66; 0,67; 0,68; 0,67; 0,69; 0,68; 0,68. Обчислити оцінку істинного вмісту марганцю при довірчій ймовірності γ=0,95.

23. Проведено вимірювання концентрації C_x невідомого розчину шляхом порівняння з розчином відомої концентрації C_0 за формулою $C_x = C_0 \frac{d_0}{d_x}$, де d_0 і d_x – товщини шарів, що однаково поглинають монохроматичне світло. У п'яти дослідах отримані результати: $\bar{d}_0 = 5{,}7;\ s_{d_0} = 0{,}15;$ $\bar{d}_x = 8{,}5;\ s_{d_x} = 0{,}18$ (у мм). Концентрація розчину C_0=2. Оцінити істинне значення виміряної концентрації з довірчою імовірністю γ=0,95.

24. Середня маса таблетки сульфадиметоксину при десятикратному зважуванні дорівнює 0,528 г. абсолютна похибка при зважуванні маси таблетки при довірчій ймовірності 0,95 $\overline{\Delta m} = 0{,}002$ г. Середній об'єм таблетки $\bar{V} = 0{,}2$ см^3. Визначити, з якою абсолютною похибкою $\bar{V}$ потрібно провести вимірювання об'єму, щоб абсолютна похибка густини сульфадиметоксину дорівнювала $\overline{\Delta\rho}$ =0,04 г/см^3.