

ЛЕКЦІЯ 17

ТОЧКОВІ ТА ІНТЕРВАЛЬНІ ОЦІНКИ ПАРАМЕТРІВ РОЗПОДІЛУ

17.1. Статистичні оцінки параметрів розподілу

Параметри досліджуваної ознаки генеральної сукупності X , з якої зроблено вибірку $x = (x_1, x_2, \dots, x_n)$ є величинами сталими, але їхні числові значення невідомі. Ці параметри оцінюються параметрами вибірки.

Нехай θ — невідомий параметр (наприклад, математичне сподівання, дисперсія тощо) досліджуваної ознаки X .

Вибірковою оцінкою параметра θ називається число θ^* , знайдене за даними вибірки x або за статистичним розподілом, залежне від x_1, x_2, \dots, x_n , і таке, що наближено дорівнює оцінюваному параметру: $\theta^* \approx \theta$.

Ця оцінка, яка визначається у даній вибірці одним числом, точкою, називається *точковою*.

Оскільки θ^* є випадковою величиною, то вона може бути зміщеною і незміщеною. Коли математичне сподівання цієї оцінки точно дорівнює оцінюваному параметру θ , тобто:

$$M(\theta^*) = \theta,$$

то оцінка θ^* називається *незміщеною*.

Якщо $M(\theta^*) \neq \theta$, то статистична оцінка θ^* називається *зміщеною* відносно параметра θ генеральної сукупності.

Різниця

$$\theta^* - \theta = \delta$$

називається *зміщенням* статистичної оцінки θ^* .

Точкова статистична оцінка θ^* називається *спроможною*, якщо, у разі необмеженого збільшення обсягу вибірки, θ^* наближається до оцінюваного параметра θ :

$$\lim_{n \rightarrow +\infty} P\{|\theta^* - \theta| < \delta\} = 1.$$

Оцінюваний параметр може мати кілька точкових незміщених статистичних оцінок. Точкова статистична оцінка називається *ефективною*, коли при заданому обсязі вибірки вона має мінімальну дисперсію.

Для невідомого математичного сподівання $M(X)$ досліджуваної ознаки (генеральної сукупності) X зміщеною точковою оцінкою є *вибіркова середня*, яка позначається \bar{x}_B і обчислюється за формулою (17.1):

$$\bar{x}_B = \frac{1}{n} \cdot \sum_{i=1}^k (x_i \cdot n_i). \quad (17.1)$$

Вона задовольняє всі три перелічені умови. Для невідомої дисперсії генеральної сукупності $D(X)$ спроможною оцінкою є *вибіркова дисперсія* D_B , обчислювана за формулою (17.2):

$$D_B = \frac{1}{n} \cdot \sum_{i=1}^k ((x_i - \bar{x}_B)^2 \cdot n_i). \quad (17.2)$$

Проте ця оцінка є зміщеною. Тому для оцінки невідомої дисперсії застосовують так звану *виправлену дисперсію*, яку позначають s^2 і обчислюють за формулою (17.3):

$$s^2 = \frac{n}{n-1} \cdot D_B = \frac{1}{n-1} \cdot \sum_{i=1}^k ((x_i - \bar{x}_B)^2 \cdot n_i). \quad (17.3)$$

Оцінкою для середнього квадратичного відхилення є *виправлене вибіркве середнє квадратичне відхилення* (17.4):

$$\sigma_B = s. \quad (17.4)$$

Для неперервної ознаки X , для якої побудовано інтервальний статистичний розподіл, точкові оцінки невідомих математичного сподівання і дисперсії також знаходять за формулами (17.2) і (17.3), в які замість варіант x_i підставляють середини (центри) x_i^* інтервалів.

Приклад 17.1. Знайти точкові оцінки математичного сподівання, дисперсії і середнього квадратичного відхилення добової виручки авіакомпанії за даними вибірки з прикладу 16.1 лекції 16.

Розв'язання. За дискретним статистичним розподілом (табл. 16.2 лекції 16) обчислюємо вибіркву середню \bar{x}_B , виправлену вибіркву дисперсію s^2 і виправлене вибіркве середнє квадратичне відхилення σ_B , використовуючи формули (17.1)-(17.4):

$$\begin{aligned} \bar{x}_B &= \frac{1}{40} \cdot (0,81 \cdot 3 + 0,82 \cdot 4 + 0,85 \cdot 6 + 0,87 \cdot 9 + 0,90 \cdot 8 + 0,94 \cdot 6 + \\ &\quad 0,97 \cdot 2 + 0,99 \cdot 2) = 0,885; \\ s^2 &= \frac{1}{39} \cdot ((0,81 - 0,855)^2 \cdot 3 + (0,82 - 0,855)^2 \cdot 4 + (0,85 - 0,855)^2 \cdot 6 + \\ &\quad (0,87 - 0,855)^2 \cdot 9 + (0,90 - 0,855)^2 \cdot 8 + (0,94 - 0,855)^2 \cdot 6 + \\ &\quad (0,97 - 0,855)^2 \cdot 2 + (0,99 - 0,855)^2 \cdot 2) = 0,0025; \\ \sigma_B &= 0,05. \end{aligned}$$

Знайдемо оцінки \bar{x}_B , s^2 і σ_B за інтервальним статистичним розподілом (табл. 16.3 лекції 16), замінюючи у формулах (17.2) і (17.3) x_i центрами частинних інтервалів x_i^* . Для цього перейдемо спочатку від інтервального статистичного розподілу до дискретного, який набирає такого вигляду (табл. 17.1):

Таблиця 17.1. Дискретний варіативний ряд

x_i^*	0,825	0,855	0,885	0,915	0,945	0,975
n_i^*	7	10	9	4	6	4

Тоді використовуючи формули (17.1)-(17.4) отримаємо:

$$\begin{aligned} \bar{x}_B &= \frac{1}{40} \cdot (0,825 \cdot 7 + 0,855 \cdot 10 + 0,885 \cdot 9 + 0,915 \cdot 4 + 0,945 \cdot 6 + \\ &\quad 0,975 \cdot 4) = 0,888; \\ s^2 &= \frac{1}{39} \cdot ((0,825 - 0,888)^2 \cdot 7 + (0,855 - 0,888)^2 \cdot 10 + (0,885 - 0,888)^2 \cdot 9 + \\ &\quad (0,915 - 0,888)^2 \cdot 4 + (0,945 - 0,888)^2 \cdot 6 + (0,975 - 0,888)^2 \cdot 4) = \\ &\quad 0,0023; \\ \sigma_B &= 0,048. \end{aligned}$$

Відповідь. $\bar{x}_B = 0,888$; $s^2 = 0,0023$; $\sigma_B = 0,048$.

17.2. Метод моментів

Метод моментів ґрунтується на тому, що невідомі параметри теоретичного розподілу (розподілу генеральної сукупності) визначаються із рівнянь, які добуваються прирівнюванням важливіших числових характеристик (моментів) теоретичного розподілу відповідним числовим характеристикам емпіричного розподілу.

Так, нехай заданий, наприклад, вид теоретичного розподілу, який визначається невідомим параметром. Для знаходження одного параметра необхідне одне рівняння відносно даного параметра. Для цього використовується момент 1-го порядку (математичне сподівання) теоретичного розподілу і відповідна числова характеристика емпіричного розподілу – вибіркове середнє.

Знаходимо математичне сподівання:

$$M(X) = \int_{-\infty}^{\infty} x d(F(x, \theta)) = M(\theta)$$

і вибіркове середнє:

$$\bar{x}_B = \frac{1}{n} \cdot \sum_{i=1}^k (x_i \cdot n_i).$$

Порівнюючи їх, одержуємо рівняння для визначення оцінки θ^* невідомого параметра (17.5):

$$M(\theta) = \bar{x}_B. \quad (17.5)$$

Для знаходження оцінок двох невідомих параметрів θ_1, θ_2 звичайно беруть математичне сподівання і дисперсію теоретичного розподілу та відповідні їм числові характеристики емпіричного розподілу – вибіркове середнє і вибіркovu дисперсію. Одержують два рівняння (17.6):

$$M(\theta_1, \theta_2) = \bar{x}_B, \sigma^2(\theta_1, \theta_2) = s^2. \quad (17.6)$$

В деяких підручниках, зокрема Гмурмана та Кремера, в системі (17.6) дисперсію теоретичного розподілу (генеральну дисперсію) оцінюють за допомогою вибіркової дисперсії D_B .

Розв'язуючи цю систему, знаходять відповідні оцінки θ_1^*, θ_2^* . Оцінки методу моментів звичайно є **слушними**, однак за ефективністю вони не є «найкращими». Тим не менш, метод моментів часто використовується на практиці, оскільки приводить до порівняно простих обчислень.

Приклад 17.2. Генеральна сукупність має експоненціальний розподіл із щільністю розподілу $f(x, \lambda) = \lambda \cdot e^{-\lambda \cdot x}$ ($x \geq 0$), де – невідомий параметр. Вибіркові значення x_1, x_2, \dots, x_n , узяті із однієї і тієї ж генеральної сукупності. Знайти методом моментів оцінку $\tilde{\lambda}$ невідомого параметра розподілу.

Розв'язання. Визначимо математичне сподівання експоненціального розподілу:

$$M(\lambda) = \lambda \cdot \int_0^{\infty} (x \cdot e^{-\lambda \cdot x}) dx = \frac{1}{\lambda}.$$

Далі по вибірці знаходимо вибіркове середнє емпіричного розподілу:

$$\bar{x}_B = \frac{1}{n} \cdot \sum_{i=1}^k (x_i \cdot n_i).$$

Із рівняння

$$M(\lambda) = \bar{x}_B = \frac{1}{\lambda}$$

знаходимо оцінку параметра:

$$\lambda^* = \frac{1}{\bar{x}_B}.$$

Відповідь. $\lambda^* = \frac{1}{\bar{x}_B}.$

Приклад 17.3. Вибірка x_1, x_2, \dots, x_n , одержана із генеральної сукупності з нормальним розподілом із щільністю розподілу:

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x-a)^2}{2 \cdot \sigma^2}}, \quad (-\infty < x < \infty),$$

де a і σ – невідомі параметри розподілу. Знайдемо методом моментів оцінки a^* і σ^* невідомих параметрів a і σ .

Розв’язання. Визначимо математичне сподівання і дисперсію розподілу:

$$a = \int_{-\infty}^{\infty} (x \cdot f(x)) dx, \quad \sigma^2 = \int_{-\infty}^{\infty} ((x-a)^2 \cdot f(x)) dx.$$

За даними вибірки знаходимо вибіркові числові характеристики \bar{x}_B, s і прирівнюємо їх до відповідних числових характеристик теоретичного розподілу:

$$a^* = \bar{x}_B, \quad \sigma^* = s,$$

що є точковими оцінками відповідних параметрів.

Відповідь. $a^* = \bar{x}_B, \quad \sigma^* = s.$

17.3. Метод максимальної правдоподібності

Одним із найбільш універсальних методів одержання оцінок параметрів розподілів генеральної сукупності є метод *максимальної правдоподібності*.

Оснoву методa склaдає функція правдоподібності, яка виражає щільність ймовірності (ймовірність) сумісної появи результатів вибірки x_1, x_2, \dots, x_n :

$$L(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \dots \cdot f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) \quad (17.7)$$

або:

$$L(x_1, x_2, \dots, x_n, \theta) = p(x_1, \theta) \cdot p(x_2, \theta) \cdot \dots \cdot p(x_n, \theta) = \prod_{i=1}^n p(x_i, \theta). \quad (17.8)$$

Згідно з методом максимальної правдоподібності за оцінку невідомого параметра приймається таке його значення, яке максимізує функцію $L(\theta)$. Величина, при якій функція правдоподібності досягає максимального значення, називається *оцінкою максимальної правдоподібності*.

Природність такого підходу до визначення статистичних оцінок впливає із смислу функції правдоподібності, яка при кожному

фіксованому значенні є *мірою правдоподібності* одержання вибірки. Оцінка є такою, що вибірка, яка одержана у результаті спостережень, є найбільш вірогідною. Знаходження оцінки спрощується, якщо максимізувати не саму функцію $L(\theta)$, а $\ln(L(\theta))$, оскільки максимум обох функцій досягається при одному і тому ж значенні.

Для одержання оцінки максимальної правдоподібності треба розв'язати рівняння:

$$\frac{d(\ln(L(\theta)))}{d\theta} = 0.$$

Якщо потрібно оцінити не один, а декілька параметрів $\theta_1, \theta_2, \dots, \theta_m$, то оцінка максимальної правдоподібності цих параметрів знаходиться із системи рівнянь:

правдоподібності полягає у тому, що для широкого класу розподілів він приводить до оцінок, які є слушними, асимптотично ефективними, мають асимптотично нормальний розподіл і, якщо для параметра існує ефективна оцінка, то рівняння правдоподібності має єдиний розв'язок, який співпадає з нею. Однак оцінка максимальної правдоподібності може

$$\frac{\partial(\ln(L(x_1, x_2, \dots, x_n, \lambda)))}{\partial \theta_i} \quad (i = 1, \dots, m).$$

Перевага методу максимальної виявиться зсуненою.

Приклад 17.4. Генеральна сукупність має розподіл Пуассона.

$$P_n(X = x_i) = \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda},$$

де n – кількість випробувань у кожній серії, x_i – кількість появ події у i -й серії ($i = 1, \dots, n$). Знайти оцінку невідомого параметра по вибірці.

Розв'язання. Складемо функцію правдоподібності.

$$\begin{aligned} L(x_1, x_2, \dots, x_n, \lambda) &= p(x_1, \lambda) \cdot p(x_2, \lambda) \cdot \dots \cdot p(x_n, \lambda) = \\ &= \left(\frac{\lambda^{x_1}}{x_1!} \cdot e^{-\lambda} \right) \cdot \dots \cdot \left(\frac{\lambda^{x_n}}{x_n!} \cdot e^{-\lambda} \right) = \frac{\lambda^{\sum x_i}}{(x_1!) \cdot \dots \cdot (x_n!)} \cdot e^{-n \cdot \lambda}. \end{aligned}$$

Визначимо логарифм цієї функції

$$\ln(L(x_1, x_2, \dots, x_n, \lambda)) = \sum (x_i \cdot \ln(\lambda)) - n \cdot \lambda - \ln((x_1!) \cdot \dots \cdot (x_n!))$$

Прирівнюючи похідну цієї функції по до нуля, одержуємо рівняння правдоподібності

$$\frac{d(\ln(L(x_1, x_2, \dots, x_n, \lambda)))}{d\lambda} = \frac{\sum x_i}{\lambda} - n = 0.$$

Розв'язуючи це рівняння відносно λ , знаходимо:

$$\lambda^* = \frac{\sum x_i}{n} = \bar{x}_B.$$

Оскільки

$$\frac{d^2(\ln(L(x_1, x_2, \dots, x_n, \lambda)))}{d\lambda^2} = -\frac{\sum x_i}{\lambda^2} < 0,$$

то $\lambda^* = \bar{x}_B$ є точкою максимуму функції $\ln(L(x_1, x_2, \dots, x_n, \lambda))$.

З вищенаведеного одержуємо, що є оцінка максимальної правдоподібності параметра розподілу Пуассона.

Відповідь. $\lambda^* = \bar{x}_B$.

Приклад 17.5. Вибірка x_1, x_2, \dots, x_n одержана із генеральної сукупності, розподіленої за нормальним законом з відповідними параметрами. Знайдемо оцінки параметрів і методом максимальної правдоподібності.

Розв'язання. Записуємо функцію правдоподібності:

$$L(x_1, x_2, \dots, x_n, a, \sigma^2) = \prod_{i=1}^n \left(\frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x_i - a)^2}{2 \cdot \sigma^2}} \right) = \frac{1}{(\sigma^2)^{\frac{n}{2}} \cdot (2 \cdot \pi)^{\frac{n}{2}}} \cdot e^{-\frac{\sum_{i=1}^n (x_i - a)^2}{2 \cdot \sigma^2}}.$$

Логарифмуючи, отримаємо:

$$\ln(L(a, \sigma)) = -\frac{n}{2} \cdot (\ln(2 \cdot \pi) + \ln(\sigma^2)) - \frac{1}{2 \cdot \sigma^2} \cdot \sum_{i=1}^n (x_i - a)^2.$$

Диференціюючи $\ln(L(a, \sigma))$ по a і σ^2 одержуємо систему рівнянь:

$$\begin{cases} \frac{1}{\sigma^2} \cdot \sum_{i=1}^n (x_i - a) = 0, \\ -\frac{n}{2 \cdot \sigma^2} + \left(\frac{1}{2 \cdot \sigma^4} \right) \cdot \sum_{i=1}^n (x_i - a)^2 = 0. \end{cases}$$

Звідки знаходимо оцінки:

$$a^* = \bar{x}_B, \sigma^* = s.$$

Ці оцінки співпадають з оцінками методу моментів. Вони обидві є слушні та незміщеними.

Відповідь. $a^* = \bar{x}_B, \sigma^* = s$.

17.4. Метод найменших квадратів

Сутність **методу найменших квадратів** полягає у тому, що оцінки невідомих параметрів розподілу $\theta = (\theta_1, \theta_2, \dots, \theta_m)$ визначаються із умови мінімізації суми квадратів відхилень вибірових даних x_i ($i = 1, \dots, n$) від оцінки, що визначається. Наприклад, знайдемо оцінку за методом найменших квадратів для генеральної середньої. Згідно з цим методом оцінку знаходимо з умови:

$$s(\theta) = \sum_{i=1}^n (x_i - \theta)^2 \rightarrow \min. \quad (17.9)$$

Використовуючи необхідну умову екстремуму функції, прирівнюємо до нуля похідну

$$\frac{ds}{d\theta} = -2 \cdot \sum_{i=1}^n (x_i - \theta) = 0 \text{ або } \sum_{i=1}^n x_i - n \cdot \theta = 0.$$

Звідки:

$$\theta^* = \bar{x}_B, \quad (17.10)$$

Отже, оцінка генеральної середньої є вибірова середня \bar{x}_B

Метод найменших квадратів має широке застосування у практиці статистичних досліджень, оскільки, по-перше, не потребує знання закону

розподілу вибірових даних; по-друге, для нього досить добре розроблений математичний апарат чисельної реалізації.

Метод найменших квадратів застосовується у моделях кореляційного і регресійного аналізу.

17.5. Метод мінімуму χ^2

Нехай x_1, x_2, \dots, x_n – вибірка незалежних спостережень над випадковою величиною X , розподіл якої належить класу розподілів $F(x, \theta)$, який залежить від невідомого параметра $\theta = (\theta_1, \theta_2, \dots, \theta_m)$. Вибірка одержана при деяких конкретних значеннях цього параметра.

Припустимо, множина значень X розбита на m інтервалів $\Delta_1, \dots, \Delta_m$, які не перетинаються. Позначимо через n_i число спостережень у вибірці, які потрапили у інтервал Δ_i . Якщо множина значень X скінченна, тобто величина приймає лише скінченне число значень, то можна вважати, що Δ_i – одноточкова множина. Таким чином, проведено групування результатів спостережень, у результаті чого одержано інтервальний варіаційний ряд.

Позначимо через $p_i(\theta) = F(\Delta_i, \theta)$, $i = 1, \dots, m$ ймовірність попадання значень випадкової величини X у i -й інтервал. Складемо величину χ^2 :

$$\chi^2 = \sum_{i=1}^m \frac{(n_i - n \cdot p_i(\theta))^2}{n \cdot p_i(\theta)}. \quad (17.11)$$

Формула (17.11) служить оцінкою параметра на основі даної вибірки. Оскільки ймовірності $p_i(\theta)$ є функціями вибірових значень, то і величина χ^2 є функцією вибірки.

Оцінка $\theta(x_1, x_2, \dots, x_n)$ називається *оцінкою за методом мінімуму χ^2* , якщо вона одержана мінімізацією функції χ^2 .

Основні кроки методу мінімуму χ^2

Збір даних: Нехай маємо n спостережень випадкової величини X , які поділені на k інтервалів (або класів). Позначимо частоту спостережень у кожному інтервалі як $O_i = n_i$ (спостережувана частота).

Розрахунок теоретичних частот: Нехай $n \cdot p_i(\theta)$ – це ймовірність того, що випадкова величина X потрапить у i -й інтервал відповідно до теоретичного розподілу з параметрами θ . Тоді теоретичну частоту для i -го інтервалу можна знайти як $E_i = n \cdot p_i(\theta)$.

Обчислення статистики χ^2 : Статистика χ^2 обчислюється за формулою (17.11).

Мінімізація χ^2 : Параметри θ оцінюються шляхом мінімізації функції χ^2 .

Оптимальні значення параметрів – це ті, при яких χ^2 досягає мінімуму.

Для мінімізації функції χ^2 , зокрема, можна скористатися алгоритмом градієнтного спуску або функцією оптимізації з бібліотеки Python, такої як SciPy (рис. 17.1).

Метод мінімуму χ^2 часто застосовується у задачах статистичного моделювання, аналізу даних та оцінювання параметрів розподілів. Він дозволяє знизити вплив випадкових похибок і досягти більш точних результатів.

Приклад 17.6. Нехай ми маємо емпіричні дані, які поділені на k інтервалів. Необхідно оцінити параметри θ для нормального розподілу $N(a, \sigma^2)$.

Розв'язання. Наведемо поетапний хід розв'язку:

Збір даних: Припустимо, що ми маємо $n = 100$ спостережень і $k = 10$ інтервалів. Спостережувані частоти для кожного інтервалу такі:

[8; 12; 15; 14; 10; 9; 12; 8; 6; 6].

Розрахунок теоретичних частот: Припустимо, що початкові оцінки параметрів нормального розподілу a та σ дорівнюють 0 і 1 відповідно. Визначаємо теоретичні частоти для кожного інтервалу.

Обчислення статистики χ^2 : Обчислюємо $\chi^2(\theta)$ за формулою (17.11).

Мінімізація χ^2 : Мінімізуємо функцію $\chi^2(\theta)$ відносно параметрів a та σ для отримання оптимальних значень параметрів.

```
python Копировать код  
  
import numpy as np  
from scipy.optimize import minimize  
from scipy.stats import norm  
  
# Спостережувані частоти  
O = np.array([8, 12, 15, 14, 10, 9, 12, 8, 6, 6])  
  
# Границі інтервалів (наприклад)  
a = np.linspace(-3, 3, 11)  
  
# Функція для обчислення теоретичних частот  
def expected_frequencies(mu, sigma):  
    E = []  
    for i in range(len(a) - 1):  
        E.append(100 * (norm.cdf(a[i+1], loc=mu, scale=sigma) - norm.cdf(a[i], loc=mu, scale=sigma)))  
    return np.array(E)  
  
# Функція для обчислення  $\chi^2$   
def chi2(params):  
    mu, sigma = params  
    E = expected_frequencies(mu, sigma)  
    return np.sum((O - E) ** 2 / E)  
  
# Початкові оцінки параметрів  
initial_params = [0, 1]  
  
# Мінімізація  
result = minimize(chi2, initial_params, method='BFGS')  
mu_opt, sigma_opt = result.x  
  
print(f"Оцінка параметра  $\mu$ : {mu_opt}")  
print(f"Оцінка параметра  $\sigma$ : {sigma_opt}")
```

Рис. 17.1. Приклад мінімізації функції $\chi^2(\theta)$ в Python

Відповідь. Оптимальні значення a^* та σ^* .

17.6. Інтервальні статистичні оцінки.

Довірча ймовірність та довірчий інтервал

Точкові оцінки параметрів розподілу не дають можливості зробити висновки про їхню точність та надійність, оскільки є по суті випадковими величинами. Щоб мати уявлення про точність та надійність оцінки, у математичній статистиці користуються довірчими ймовірностями та довірчими інтервалами.

Інтервал $(-\theta^*; \theta^*)$ називається **довірчим** для оцінюваного параметра θ з довірчою ймовірністю (надійністю) γ , якщо він покриває цей параметр θ з ймовірністю, не меншою за γ , тобто

$$P\{-\theta^* < \theta < \theta^*\} \geq \gamma.$$

Межі $-\theta^*$ і θ^* довірчого інтервалу називаються довірчими межами для оцінюваного параметра θ . Статистична оцінка, що визначається двома числами — кінцями інтервалів, називається інтервальною. **Надійність γ** вибирається, як правило, достатньо великою, наприклад, 0,9; 0,95 або 0,99.

17.7. Схема побудови довірчого інтервалу для невідомого параметра σ нормально розподіленої ознаки X при відомому параметрі σ

Довірчий інтервал матиме вигляд (17.12):

$$\left(\bar{x}_B - t \cdot \frac{\sigma}{\sqrt{n}}; \bar{x}_B + t \cdot \frac{\sigma}{\sqrt{n}} \right), \quad (17.12)$$

де \bar{x}_B — вибіркова середня; n — об'єм вибірки; t — розв'язок рівняння $\Phi(t) = \frac{\gamma}{2}$ при заданій надійності γ . Для функції Лапласа $\Phi(x)$ складено таблицю значень (дод. 2 Гмурмана), за допомогою якої за заданим значенням $\frac{\gamma}{2}$ функції знаходять значення аргументу t . Величина $\delta = t \cdot \frac{\sigma}{\sqrt{n}}$ називається **похибкою**.

Приклад 17.7. Маємо такі дані про розміри виручки (у тис. грн) на 30 випадково вибраних рейсах:

42 24 49 67 45 27 39 21 58 40
28 78 44 66 20 62 70 81 7 68
94 76 63 88 65 14 46 20 72 91

Побудувати інтервальний статистичний розподіл із довжиною кроку $h = 20$ (тис. грн). Припускаючи, що виручка розподілена за нормальним законом, знайти з надійністю $\gamma = 0,95$ довірчий інтервал для невідомого математичного сподівання, якщо середнє квадратичне відхилення $\sigma = 25$ (тис. грн).

Розв'язання. Інтервальний статистичний розподіл буде таким (табл. 17.2):
Таблиця 17.2. Інтервальний варіативний ряд

$[x_i; x_{i+1})$	[7; 27)	[27; 47)	[47; 67)	[67; 87)	[87; 107)
n_i	6	8	6	7	3

Для визначення \bar{x}_B необхідно побудувати дискретний статистичний розподіл, що має такий вигляд (табл.13.3):

Таблиця 17.3. Дискретний варіативний ряд

x_i^*	17	37	57	77	97
n_i	6	8	6	7	3

де x_i^* — середини відповідних інтервалів.

Тоді за формулою (17.1)

$$\bar{x}_B = \frac{1}{30} \cdot (17 \cdot 6 + 37 \cdot 8 + 57 \cdot 6 + 77 \cdot 7 + 97 \cdot 3) = 52,3.$$

Для побудови довірчого інтервалу із заданою надійністю $\gamma = 0,95$ необхідно знайти t : $\Phi(t) = \frac{\gamma}{2} = 0,475$. Звідси $t = 1,96$ (дод. 2 Гмурмана).

Знаходимо межі інтервалу:

$$\begin{aligned}\bar{x}_B - t \cdot \frac{\sigma}{\sqrt{n}} &= 52,3 - 1,96 \cdot \frac{25}{\sqrt{30}} = 43,4, \\ \bar{x}_B + t \cdot \frac{\sigma}{\sqrt{n}} &= 52,3 + 1,96 \cdot \frac{25}{\sqrt{30}} = 61,2.\end{aligned}$$

Відповідь. З ймовірністю, не меншою за 0,95, середнє значення виручки на даному рейсі міститься в межах від 43,4 до 61,2 тис. грн.

Приклад 17.8. Визначити об'єм вибірки n нормально розподіленої випадкової величини, за якого похибка $\delta = 0,01$ при обчисленні математичного сподівання гарантується з ймовірністю 0,999, якщо середнє квадратичне відхилення $\sigma = 5$.

Розв'язання. Оскільки $\delta = \frac{t \cdot \sigma}{\sqrt{n}}$, то звідси $n = \frac{t^2 \cdot \sigma^2}{\delta^2}$. Значення t знаходимо з рівняння $\Phi(t) = \frac{\gamma}{2} = \frac{0,999}{2} = 0,4995$. За таблицею функції Лапласа (дод. 2 Гмурмана) знаходимо $t = 3,4$, отже, $n = 2890000$.

Відповідь. $n = 2890000$.

17.8. Схема побудови довірчого інтервалу для невідомого параметра σ нормально розподіленої ознаки X при відомому параметрі μ

Оскільки при розв'язуванні задач математичної статистики ми, як правило, маємо в розпорядженні лише вибірку, тобто параметр σ невідомий, то для знаходження наближених меж довірчого інтервалу замість σ використовується його вибіркова точкова оцінка s . При великих обсягах вибірки така заміна приводить до цілком прийнятних похибок.

При малих вибірках для оцінювання математичного сподівання якщо невідоме значення середнього квадратичного відхилення, застосовується випадкова величина t , що має розподіл Стюдента з $k = n - 1$ ступенями свободи та рівнем значущості $\alpha = 1 - \gamma$ (дод. 6 Гмурмана).

При цьому для кожного γ ($0 < \gamma < 1$) існує єдине значення t_γ , яке визначається за таблицею (дод. 3 Гмурмана) з для надійності γ та об'єму вибірки n .

Обчисливши за даним статистичним розподілом \bar{x}_B , s^2 і визначивши значення t_γ , будемо довірчий інтервал (17.13):

$$\left(\bar{x}_B - t_\gamma \cdot \frac{s}{\sqrt{n}}; \bar{x}_B + t_\gamma \cdot \frac{s}{\sqrt{n}} \right). \quad (17.13)$$

Приклад 17.9. Побудувати довірчий інтервал з надійністю $\gamma = 0,99$ для математичного сподівання максимальної швидкості (м/с) літака, припускаючи, що вона має нормальний розподіл, за даними 20 вимірювань:

504,5 485,2 512,0 497,1 502,4 488,3 491,9 489,3 502,7 509,2
514,3 502,3 497,8 511,3 515,7 506,3 499,2 497,4 485,2 507,8

Розв'язання. Для побудови довірчого інтервалу знайдемо середнє вибіркове і виправлену дисперсію.

$$\begin{aligned} \bar{x}_B &= \frac{1}{20} \cdot (504,5 + 485,2 + 512,0 + 497,1 + 502,4 + 488,3 + 491,9 + \\ &489,3 + 502,7 + 509,2 + 514,3 + 502,3 + 497,8 + 511,3 + 515,7 + \\ &506,3 + 499,2 + 497,4 + 485,2 + 507,8) \approx 501; \\ s^2 &= \frac{1}{19} \cdot ((504,5 - 501)^2 + (485,2 - 501)^2 + (512,0 - 501)^2 + \\ &(497,1 - 501)^2 + (502,4 - 501)^2 + (488,3 - 501)^2 + (491,9 - 501)^2 + \\ &(489,3 - 501)^2 + (502,7 - 501)^2 + (509,2 - 501)^2 + (514,3 - 501)^2 + \\ &(502,3 - 501)^2 + (497,8 - 501)^2 + (511,3 - 501)^2 + (515,7 - 501)^2 + \\ &(506,3 - 501)^2 + (499,2 - 501)^2 + (497,4 - 501)^2 + (485,2 - 501)^2 + \\ &(507,8 - 501)^2) \approx 89,6. \end{aligned}$$

Відповідно до таблиці (дод. 3 Гмурмана) за даною надійністю $\gamma = 0,99$ і об'ємом вибірки $n = 20$ знаходимо значення $t_\gamma = 2,861$. Обчислимо межі довірчого інтервалу:

$$\begin{aligned} \bar{x}_B - t_\gamma \cdot \frac{s}{\sqrt{n}} &= 501 - \frac{2,861 \cdot \sqrt{89,6}}{\sqrt{20}} = 495; \\ \bar{x}_B + t_\gamma \cdot \frac{s}{\sqrt{n}} &= 501 + \frac{2,861 \cdot \sqrt{89,6}}{\sqrt{20}} = 507. \end{aligned}$$

Відповідь. З ймовірністю, не меншою за $\gamma = 0,99$, можна стверджувати, що математичне сподівання максимальної швидкості даного літака міститься в інтервалі (495; 507).

17.9. Схема 1 побудови довірчого інтервалу для невідомого параметра σ нормально розподіленої ознаки X при відомому параметрі s

При невідомому математичному сподіванні довірчий інтервал із надійністю γ для дисперсії $D(X)$ нормального розподілу має вигляд (17.14):

$$\left(\frac{k \cdot s^2}{\chi^2(k; \alpha_1)}; \frac{k \cdot s^2}{\chi^2(k; \alpha_2)} \right), \quad (17.14)$$

а для середнього квадратичного відхилення дістаємо (17.15):

$$\left(\frac{\sqrt{k} \cdot s}{\chi(k; \alpha_1)}; \frac{\sqrt{k} \cdot s}{\chi(k; \alpha_2)} \right), \quad (17.15)$$

де α_1, α_2 — рівні значущості, а $k = n - 1$ — ступінь свободи випадкової величини $\chi^2(k; \alpha)$.

Додатні числа $\chi^2(k; \alpha_1)$ та $\chi^2(k; \alpha_2)$ визначаються з рівностей (17.16):

$$F(\chi^2(k; \alpha_1)) = \alpha_1 = \frac{1-\gamma}{2}; F(\chi^2(k; \alpha_2)) = \alpha_2 = \frac{1+\gamma}{2}, \quad (17.16)$$

де $F(\chi^2(k; \alpha))$ — функція розподілу випадкової величини $\chi^2(k; \alpha)$ з k ступенями свободи і рівнем значущості $\alpha \in [0,1]$.

Відповідне значення $\chi^2(k; \alpha)$ шукається як значення аргументу, при якому функція $F(\chi^2(k; \alpha))$ набуде значення α (дод. 5 Гмурмана) або за таблицею квантилів відповідного розподілу (рис. 17.2).

	0,01	0,025	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,975	0,99
1	0,0002	0,0010	0,0039	0,0158	0,0642	0,1485	0,2750	0,4549	0,7083	1,0742	1,6424	2,7055	3,8415	5,0239	6,6349
2	0,0201	0,0506	0,1026	0,2107	0,4463	0,7133	1,0217	1,3863	1,8326	2,4079	3,2189	4,6052	5,9915	7,3778	9,2103
3	0,1148	0,2158	0,3518	0,5844	1,0052	1,4237	1,8692	2,3660	2,9462	3,6649	4,6416	6,2514	7,8147	9,3484	11,3449
4	0,2971	0,4844	0,7107	1,0636	1,6488	2,1947	2,7528	3,3567	4,0446	4,8784	5,9886	7,7794	9,4877	11,1433	13,2767
5	0,5543	0,8312	1,1455	1,6103	2,3425	2,9999	3,6555	4,3515	5,1319	6,0644	7,2893	9,2364	11,0705	12,8325	15,0863
6	0,8721	1,2373	1,6354	2,2041	3,0701	3,8276	4,5702	5,3481	6,2108	7,2311	8,5581	10,6446	12,5916	14,4494	16,8119
7	1,2390	1,6899	2,1673	2,8331	3,8223	4,6713	5,4932	6,3458	7,2832	8,3834	9,8032	12,0170	14,0671	16,0128	18,4753
8	1,6465	2,1797	2,7326	3,4895	4,5936	5,5274	6,4226	7,3441	8,3505	9,5245	11,0301	13,3616	15,5073	17,5345	20,0902
9	2,0879	2,7004	3,3251	4,1682	5,3801	6,3933	7,3570	8,3428	9,4136	10,6564	12,2421	14,6837	16,9190	19,0228	21,6660
10	2,5582	3,2470	3,9403	4,8652	6,1791	7,2672	8,2955	9,3418	10,4732	11,7807	13,4420	15,9872	18,3070	20,4832	23,2093
11	3,0535	3,8157	4,5748	5,5778	6,9887	8,1479	9,2373	10,3410	11,5298	12,8987	14,6314	17,2750	19,6751	21,9200	24,7250
12	3,5706	4,4038	5,2260	6,3038	7,8073	9,0343	10,1820	11,3403	12,5838	14,0111	15,8120	18,5493	21,0261	23,3367	26,2170
13	4,1069	5,0088	5,8919	7,0415	8,6339	9,9257	11,1291	12,3398	13,6356	15,1187	16,9848	19,8119	22,3620	24,7356	27,6882
14	4,6604	5,6287	6,5706	7,7895	9,4673	10,8215	12,0785	13,3393	14,6853	16,2221	18,1508	21,0641	23,6848	26,1189	29,1412
15	5,2293	6,2621	7,2609	8,5468	10,3070	11,7212	13,0297	14,3389	15,7332	17,3217	19,3107	22,3071	24,9958	27,4884	30,5779
16	5,8122	6,9077	7,9616	9,3122	11,1521	12,6243	13,9827	15,3385	16,7795	18,4179	20,4651	23,5418	26,2962	28,8454	31,9999
17	6,4078	7,5642	8,6718	10,0852	12,0023	13,5307	14,9373	16,3382	17,8244	19,5110	21,6146	24,7690	27,5871	30,1910	33,4087
18	7,0149	8,2307	9,3905	10,8649	12,8570	14,4399	15,8932	17,3379	18,8679	20,6014	22,7595	25,9894	28,8693	31,5264	34,8053
19	7,6327	8,9065	10,1170	11,6509	13,7158	15,3517	16,8504	18,3377	19,9102	21,6891	23,9004	27,2036	30,1435	32,8523	36,1909
20	8,2604	9,5908	10,8508	12,4426	14,5784	16,2659	17,8088	19,3374	20,9514	22,7745	25,0375	28,4120	31,4104	34,1696	37,5662
21	8,8972	10,2829	11,5913	13,2396	15,4446	17,1823	18,7683	20,3372	21,9915	23,8578	26,1711	29,6151	32,6706	35,4789	38,9322
22	9,5425	10,9823	12,3380	14,0415	16,3140	18,1007	19,7288	21,3370	23,0307	24,9390	27,3015	30,8133	33,9244	36,7807	40,2894
23	10,1957	11,6886	13,0905	14,8480	17,1865	19,0211	20,6902	22,3369	24,0689	26,0184	28,4288	32,0069	35,1725	38,0756	41,6384
24	10,8564	12,4012	13,8484	15,6587	18,0618	19,9432	21,6525	23,3367	25,1063	27,0960	29,5533	33,1962	36,4150	39,3641	42,9798

Рис. 17.2. Фрагмент таблиці квантилів розподілу $\chi^2(k; \alpha)$

Слід зазначити, що в даній таблиці, випадкова величина, що має розподіл типу χ^2 з k ступенями свободи і рівнем значущості α , не перевищує значення $\chi^2(k; \alpha)$ з ймовірністю α і перевищує його з ймовірністю $1 - \alpha$. Даний факт впливає прямо з означення квантилів. Це означає, що наприклад рівню значущості $\alpha = 0,1$ та степеню свободи $k = 10$ відповідає $\chi^2(k; 1 - \alpha) = \chi^2(10; 0,9) = 15,9872$.

На рис. 17.3. наведено приклад розрахунку квантилів розподілу $\chi^2(k; \alpha)$ за допомогою відповідної функція в програмному забезпеченні для статистичного аналізу (у Python). Деякі розраховані квантелі представлено на рис. 17.4.

```
python
Копировать код

import scipy.stats as stats

# Задати рівні значущості та ступені свободи
alphas = [0.10, 0.05, 0.01]
degrees_of_freedom = [1, 5, 10, 20, 30, 40, 50]

# Розрахувати квантілі
quantiles = {}
for df in degrees_of_freedom:
    quantiles[df] = {}
    for alpha in alphas:
        quantile = stats.chi2.ppf(1 - alpha, df)
        quantiles[df][alpha] = quantile

# Вивести результати
for df in degrees_of_freedom:
    print(f"Ступені свободи: {df}")
    for alpha in alphas:
        print(f"Квантиль для рівня значущості {alpha}: {quantiles[df][alpha]:.4f}")
```

Рис. 17.3. Приклад розрахунку квантилів розподілу $\chi^2(k; \alpha)$ у Python

	alpha=0.1	alpha=0.05	alpha=0.01
df			
1	2.705544	3.841459	6.634897
2	4.605170	5.991465	9.210340
3	6.251389	7.814728	11.344867
4	7.779440	9.487729	13.276704
5	9.236357	11.070498	15.086272
6	10.644641	12.591587	16.811894
7	12.017036	14.067140	18.475307
8	13.361566	15.507313	20.090235
9	14.683657	16.918978	21.665994
10	15.987179	18.307038	23.209251
11	17.275013	19.675138	24.725025
12	18.549348	21.026070	26.217000
13	19.811929	22.362032	27.688250
14	21.064140	23.684791	29.141238
15	22.307130	24.995790	30.577914
16	23.541856	26.296228	31.999927
17	24.769037	27.587112	33.408664
18	25.989420	28.869299	34.805306
19	27.203575	30.143527	36.190869

Рис. 17.4. Отримані результати (квантелі)

Приклад 17.10. Під час вибіркового аналізу максимального завантаження літака при певній кількості пального, дістали такі дані, т:

78 80 85 88 89 90 88 93
88 85 93 89 80 90 85 89

Припускаючи, що максимальне завантаження літака має нормальний розподіл, знайти довірчий інтервал для дисперсії та середнього квадратичного відхилення з надійністю $\gamma = 0,95$.

Розв'язання. Для побудови довірчих інтервалів необхідно знайти s^2 . Для цього обчислимо значення \bar{x}_B .

$$\bar{x}_B = \frac{1}{16} \cdot (78 \cdot 1 + 80 \cdot 2 + 85 \cdot 3 + 88 \cdot 3 + 89 \cdot 3 + 90 \cdot 2 + 92 \cdot 1 + 93 \cdot 1) = 86,8,$$

$$s^2 = \frac{1}{15} \cdot ((78 - 86,8)^2 \cdot 1 + (80 - 86,8)^2 \cdot 2 + (85 - 86,8)^2 \cdot 3 + (88 - 86,8)^2 \cdot 3 + (89 - 86,8)^2 \cdot 3 + (90 - 86,8)^2 \cdot 2 + (92 - 86,8)^2 \cdot 1 + (93 - 86,8)^2 \cdot 1) = 19.$$

За формулами (17.16) дістаємо:

$$F(\chi^2(15, \alpha_1)) = \frac{1 - 0,95}{2} = 0,025; F(\chi^2(15, \alpha_2)) = \frac{1 + 0,95}{2} = 0,975.$$

За таблицею значень величини $\chi^2(k, \alpha)$ (дод. 5 Гмурмана) з $k = n - 1 = 15$ та $\alpha_1 = 0,025$, $\alpha_2 = 0,975$ знаходимо $\chi_1^2(0,025; 15) \approx 27,5$, $\chi_2^2(0,975; 15) \approx 6,26$ і за формулами (17.14) та (17.15) дістаємо довірчі інтервали відповідно для дисперсії та середнього квадратичного відхилення (10,4; 45,5) і (3,2; 6,7).

Відповідь. (10,4; 45,5) і (3,2; 6,7).

17.10. Схема 2 побудови довірчого інтервалу для невідомого параметра σ нормально розподіленої ознаки X при відомому параметрі s

Інтервальною оцінкою з надійністю γ середнього квадратичного відхилення σ нормально розподіленої кількісної ознаки X по виправленому середньому квадратичному відхиленню s є довірчий інтервал (17.17):

$$\begin{aligned} s \cdot (1 - q) < \sigma < s \cdot (1 + q) \text{ при } q < 1, \\ 0 < \sigma < s \cdot (1 + q) \text{ при } q > 1, \end{aligned} \quad (17.17)$$

де q знаходиться за таблицею в дод. 4 Гмурмана за заданими n та γ .

Приклад 17.11. Під час вибіркового аналізу максимального завантаження літака при певній кількості пального, дістали такі дані, т:

78 80 85 88 89 90 88 93
88 85 93 89 80 90 85 89

Припускаючи, що максимальне завантаження літака має нормальний розподіл, знайти довірчий інтервал для середнього квадратичного відхилення з надійністю $\gamma = 0,95$.

Розв'язання. У прикладі 17.10. наведено обчислення s . З дод. 4 Гмурмана отримуємо $q = 0,44$. Згідно (17.17) довірчий інтервал для середнього квадратичного відхилення (2,9; 6,3).

Відповідь. (2,9; 6,3).

17.11. Схема побудови довірчого інтервалу для невідомої ймовірності p біноміально розподіленої ознаки X

Інтервальною оцінкою з надійністю γ невідомої ймовірності p біноміального розподілу по відносній частоті w є довірчий інтервал (17.18):

$$p_1 < p < p_2. \quad (17.18)$$

В нерівності (17.18) отримуємо (17.19):

$$\begin{aligned} p_1 &= \frac{n}{t^2 + n} \cdot \left(w + \frac{t^2}{2 \cdot n} - \sqrt{\frac{w \cdot (1 - w)}{n} + \left(\frac{t}{2 \cdot n} \right)^2} \right), \\ p_2 &= \frac{n}{t^2 + n} \cdot \left(w + \frac{t^2}{2 \cdot n} + \sqrt{\frac{w \cdot (1 - w)}{n} + \left(\frac{t}{2 \cdot n} \right)^2} \right), \end{aligned} \quad (17.19)$$

де n – загальна кількість випробувань, m – число появ події, w – відносна частота, t — розв'язок рівняння $\Phi(t) = \frac{\gamma}{2}$ при заданій надійності γ .

Зауваження: при великих значеннях n (порядку сотні) в нерівності (17.19) отримуємо (17.20):

$$p_1 = w - t \cdot \sqrt{\frac{w \cdot (1 - w)}{n}}, \quad p_2 = w + t \cdot \sqrt{\frac{w \cdot (1 - w)}{n}}. \quad (17.20)$$

Приклад 17.12. Проводиться 60 незалежних випробувань. В них подія A з'являється 15 разів. Отримати інтервальну оцінку з надійністю $\gamma = 0,95$ невідомої ймовірності p .

Розв'язання. Згідно умови задачі $n = 60, m = 15, w = 0,25$.

$$\Phi(t) = \frac{0,96}{2} = 0,48, \text{ тобто } t = 1,96 \text{ (див. дод. 2 Гмурмана).}$$

Згідно (17.20) довірчий інтервал для ймовірності $p - (0,16; 0,37)$.

Відповідь. $(0,16; 0,37)$