ПРАКТИЧНА РОБОТА 1

Оцінка невідомих параметрів нормального розподілу

Коли ми поширюємо уявлення про кінцеву групу осіб на інші групи або на всю сукупність, ми користуємося інформацією про вибірку. Коли лікар хоче одержати уявлення про склад і стан крові пацієнта, він проводить аналіз невеликої вибірки крові. Будь-яке значення необхідного параметра, обчислене на основі обмеженого числа дослідів, завжди міститиме елемент випадковості. Працівники охорони здоров'я постійно мають справу з інформацією, що базується на обмежених вибірках. Тому вони повинні добре уявляти собі межі надійності аналізу інформації на основі вибіркових даних.

Мета роботи:

Вивчити поняття “генеральна сукупність” і “вибірка”, навчитися обчислювати вибіркову середню, виправлену вибіркову дисперсію, виправлене середньоквадратичне відхилення, навчитися обчислювати довірчий інтервал для математичного очікування, відповідний заданій довірчій вірогідності.

1.1 Теоретичні відомості

У біологічній і медичній статистиці часто доводиться досліджувати розподіл тієї або іншої ознаки для вельми великої сукупності індивідуумів, що утворюють статистичний колектив (такою ознакою може бути, наприклад, вміст білка в зерні пшениці, вага новонародженої дитини, період коливань маятника і т.п.). Дана ознака є випадковою величиною, значення якої від індивідуума до індивідуума змінюється. Проте, для того, щоб скласти уявлення про розподіл цієї випадкової величини або про її найважливіші характеристики, немає необхідності обстежувати кожен об'єкт даної об’ємної (генеральної) сукупності, а можна обстежувати деяку вибірку досить великого об'єму для того, щоб в ній були виявлені істотні риси розподілу, що вивчається.

Статистична сукупність є безліччю об'єктів, однорідних щодо ознаки, що характеризує ці об'єкти.

Генеральною сукупністю називається сукупність, що складається зі всіх об'єктів, які можуть бути до неї віднесені. Теоретично це нескінченно велика сукупність або сукупність, що наближається до нескінченності. Число об'єктів генеральної сукупності називають її об'ємом і позначають N.

Вибірковою сукупністю або вибіркою називається безліч об'єктів, випадково відібраних з генеральної сукупності. Число об'єктів вибірки називають її об'ємом і позначають n.

Для того, щоб властивості вибірки достатньо добре відображали властивості генеральної сукупності, вибірка повинна бути здійснена випадково, тобто всі об'єкти повинні мати однакову вірогідність потрапити у вибірку.

Оскільки на практиці доводиться мати справу з обмеженою кількістю експериментальних даних, то результати спостережень і їх обробки містять більший або менший елемент випадковості.

Характеристики статистичного розподілу вибірки застосовуються для оцінки невідомих параметрів теоретичного розподілу вірогідності.

Розрізняють точкові оцінки випадкової величини (одним числом) і інтервальні (оцінювання параметра сукупності у вигляді інтервалу).

Введемо деякі поняття.

Генеральна середня $\overline{X}_Г$ — середнє арифметичне значення ознаки X1, X2 ., Xn генеральної сукупності, тобто

$$\overline{X}_Г = \frac{1}{N}\sum_{i=1}^{N} X_i .$$

Генеральна середня дорівнює математичному очікуванню випадкової величини:

$$\overline{X}_Г = \mu .$$

Вибіркова середня $\overline{X}_B$ — середнє арифметичне значення ознаки вибіркової сукупності X_1, X_2, ., Xn тобто

$$\overline{X}_B = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

Генеральна дисперсія:

$$D(x) = \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \mu\right)^2 .$$

Вибіркова дисперсія:

$$\sigma_B^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_B\right)^2 .$$

Точкові оцінки. За оцінку значення μ вимірюваної величини береться вибіркова середня:

$$\mu \approx \overline{X}_B = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

За оцінку дисперсії D береться значення виправленої вибіркової дисперсії S^2:

$$D \approx S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_B\right)^2 = \frac{n}{n-1}\sigma_B^2 .$$

Інтервальна оцінка математичного очікування (довірчий інтервал для математичного очікування випадкової величини, розподіленої по нормальному закону, при невідомому σ).

Нехай випадкова величина A має нормальний розподіл, причому невідомі μ і σ.

У ряді завдань потрібно не тільки знайти для параметра μ відповідне чисельне значення, але і оцінити його точність. Потрібно знати, до яких помилок може привести заміна параметра μ його точковою оцінкою $\overline{X}_B$ і з яким ступенем упевненості можна чекати, що ці помилки не вийдуть за відомі межі?

Такого роду завдання особливо актуальні при малому числі спостережень, коли точкова оцінка значною мірою випадкова і наближена заміна може привести до серйозних помилок.

Щоб дати уявлення про точність і надійність в математичній статистиці користуються так званими довірчим інтервалом і довірчою вірогідністю.

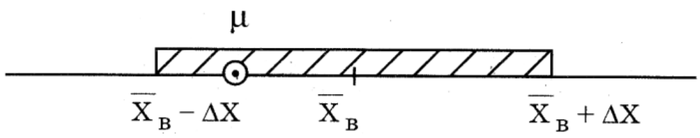
Різні вибірки дадуть різні оцінки. Нехай для параметра μ одержана точкова оцінка $\overline{X}_B$. При цьому, замінюючи μ на $\overline{X}_B$, ми здійснюємо деяку помилку.

У теорії математичної статистики показано, що із заданою вірогідністю α невідоме значення параметра μ потрапляє в певний інтервал (малюнок 1.1):

$$\left(\overline{X}_B - \Delta X, \overline{X}_B + \Delta X\right)$$

або

$$\overline{X}_B - \Delta X < \mu < \overline{X}_B + \Delta X .$$



Малюнок 1.1 - Довірчий інтервал

Вірогідність α прийнято називати довірчою вірогідністю. З такою вірогідністю ми "довіряємо" результату. Величина α вибирається самим дослідником самостійно, наприклад α = 0,95; 0,98.

Із заданою вірогідністю α довірчий інтервал накриває точку μ.

Величина X – напівширина довірчого інтервалу. Точки $\overline{X}_B + \Delta X$ и $\overline{X}_B - \Delta X$ — довірчі межі.

Величини $\overline{X}_B$ і X обчислюються на основі експериментальних даних.

Допустимий випадкова величина А підкоряється нормальному закону розподілу.

У експерименті набуте її значення:

X1, X2 ., Xn.

Якщо об'єм вибірки невеликий, (n<30), то напівширина довірчого інтервалу в цьому випадку обчислюється за формулою:

$$\Delta X = t_{\alpha,n} \cdot \frac{S}{\sqrt{n}},$$

де $t_{\alpha,n}$ – коефіцієнт Стьюдента, значення якого залежить від довірчої вірогідності α і від об'єму вибірки n. Його значення приведені в спеціальній таблиці.

Тоді довірчий інтервал для μ:

$$\left(\overline{X}_B - t_{\alpha,n} \cdot \frac{S}{\sqrt{n}}; \overline{X}_B + t_{\alpha,n} \cdot \frac{S}{\sqrt{n}}\right).$$

Таким чином, математичне очікування μ знаходиться в довірчому інтервалі:

$$\overline{X}_B - t_{\alpha,n} \cdot \frac{S}{\sqrt{n}} < \mu < \overline{X}_B + t_{\alpha,n} \cdot \frac{S}{\sqrt{n}}$$

із заданою довірчою вірогідністю α.

Чим вище ми задаємо вірогідність, тим ширше стає довірчий інтервал. І, навпаки, чим менше, тим вужче інтервал.

При збільшенні об'єму вибірки ширина інтервалу зменшується.

Приклад розрахунку ΔX. Прі вимірюванні деякої величини набуте наступного значення: X1 = 3,1; X_2 = 3,3; X3 == 3,2. З довірчою вірогідністю α = 0,95 оцінити дійсне значення вимірюваної величини.

Рішення. Обчислюємо середнє вибіркове значення

$$\overline{X}_B = \frac{3{,}1 + 3{,}3 + 3{,}2}{3} = 3{,}2.$$

Обчислюємо виправлену дисперсію

$$S^2 = \frac{(3{,}2-3{,}1)^2 + (3{,}2-3{,}3)^2 + (3{,}2-3{,}2)^2}{3-1} = 0{,}01.$$

Обчислюємо напівширину довірчого інтервалу. Значення коефіцієнта Стьюдента знаходимо в відповідній таблиці.

$$t_{0,95;3} \cong 4{,}3 \rightarrow \Delta X \cong 0{,}25 \cong 0{,}3.$$

Отже:

3,2-0,3<<3,2+0,3.

Відповідь: З довірчою вірогідністю α = 0,95 генеральне середнє вимірюваної величини знаходиться в довірчому інтервалі (2,9; 3,5).

1.2 Виконання роботи

Завдання 1. При вимірюванні періоду коливання математичного маятника отримані наступні значення:

3,0; 2,8; 3,1; 3,0; 2.9; 3,1; 2,8 с.

Оцініть довірчий інтервал для математичного очікування періоду коливань. Чи потрапляє в цей інтервал теоретичне значення періоду? Довжина маятника 0,78 м.

Завдання 2. При вимірюванні гемоглобіну у двох жінок одержані наступні дані:

1 пацієнта

Вміст гемоглобіну в крові, г/л	X1	X2	X3	X4	X5	X6	X7	X8
	128	127	126	122	128	126	129	125

2 пацієнт

Вміст гемоглобіну в крові, г/л	X1	X2	X3	X4	X5	X6	X7	X8
	85	88	90	85	86	91	88	87

Оцініть довірчий інтервал для математичного очікування концентрації гемоглобіну в крові для даних пацієнтів. Порівняйте з нормою:

$$С_{норма}=130 \pm 10 \text{ (г/л)}.$$

Завдання 3. Проаналізуйте зміну напівширини довірчого інтервалу залежно від довірчої вірогідності, що задається (при фіксованому об'ємі вибірки).

Дані вимірювання концентрації солі в розчині:

з, г/л   10; 10,2; 10,1; 10,4; 10,4; 10,3; 10,2.

Знайдіть для різних а значення ΔX і занесіть в таблицю:

α	0,6	0,8	0,9	0,95	0,99
$t_{\alpha,n}$					
ΔX					

Представте результати графічно (у вигляді малюнка 1.1).

Завдання 4. Проаналізуйте зміну напівширини довірчого інтервалу залежно від об'єму вибірки (при фіксованій довірчій вірогідності α = 0,95).

Значення випадкової величини приведені в таблиці:

Xi	100, 102, 101	99, 103	99, 101	104, 102
n	3	5	7	9
ΔX				
$\overline{X}_B$				

Результат представте графічно.

Додаток. Значення коефіцієнта Стьюдента $t_{\alpha,n}$ при різних значеннях α і n:

n/α	0,6	0,8	0,9	0,95	0,99
2	1,38	3,08	6,31	12,71	63,66
3	1,06	1,89	2,92	4,30	9,93
4	0,98	1,64	2,35	3,18	5,84
5	0,94	1,53	2,13	2,78	4,60

6	0,92	1,48	2,02	2,57	4,03
7	0,91	1,44	1,94	2,45	3,71
8	0,90	1,42	1,90	2,37	3,50
9	0,89	1,40	1,86	2,31	3,36

1.3 Контрольні питання

1. Що називається “генеральною сукупністю”? Вибірковою сукупністю?
2. Формули для обчислення генеральної середньої, вибіркової середньої, виправленої вибіркової дисперсії.
3. Яка величина є точковою оцінкою математичного очікування? Яка величина є точковою оцінкою дисперсії?
4. Сенс довірчого інтервалу, довірчої вірогідності.
5. Формули для їх розрахунків.