

Б. Пряха

Київський національний університет будівництва і архітектури

ЯВНІ ОЗНАЧЕННЯ ВИБІРКОВИХ ДИСПЕРСІЙ

© Пряха Б., 2007

*Обґрунтовано явні означення вибіркової дисперсії s^2 , S^2 .**Evident definitions of sample dispersions s^2 , S^2 are substantiated.*

Постановка проблеми. Для оцінювання стану навколишнього середовища часто виконують цифрове знімання місцевості. Точність визначення координат і висот точок місцевості по стереопарі цифрових знімків оцінюється вибілковими дисперсіями. Отже, важливо знати сутність характеристик розсіювання значень вибірки.

Всі означення підрозділяються на явні та неявні [1]. У теорії точності вимірювань будь-яке явне означення

- 1) розкриває сутність того, що означається;
- 2) наводить правило його визначення.

Неявні означення – це аксіоматичні означення, що мають круговий характер; у них вихідні терміни означаються один через одного. Такі означення наводять тільки правило обчислення того, що означається, а сутність його не розкривають.

Розсіювання значень сукупності незалежних вимірювань оцінюють за вибірковою дисперсією

$$S^2 = \frac{k}{k-1} s^2 = \frac{1}{k-1} \sum_{i=1}^k (x_i - \bar{x})^2, \quad (1)$$

де k – обсяг сукупності результатів вимірювань;

$$s^2 = \overline{(x - \bar{x})^2} = \frac{1}{k} \sum_{i=1}^k (x_i - \bar{x})^2 \quad (2)$$

i є вибірковою дисперсією; $s = \sqrt{s^2}$ – вибіркоче стандартне відхилення;

$$\bar{x} = \frac{1}{k} \sum_{i=1}^k x_i \quad (3)$$

– вибіркоче середнє значення [2].

Означення (3) – це явне означення, оскільки воно не є аксіоматичним, являє собою наслідок з закону великих чисел (теорема Хінчина). Величина \bar{x} є оцінка для середнього значення випадкової величини

$$\mu \equiv E(X) = \sum_x x f(x), \quad (4)$$

де $f(x)$ – функція розподілу ймовірностей [3].

Вибіркові дисперсії S^2, s^2 знаходяться залежно від відхилень $x_i - \bar{x}$, а згідно з (3) вибіркоче середнє значення \bar{x} означається через значення x вибірки.

Отже, означення (1), (2) наводять тільки правила обчислення вибіркових дисперсій, вони не розкривають їх сутність, а тому є неявними означеннями.

Аналіз останніх досліджень і публікацій. В [4] наведені явні означення дисперсій σ^2, Σ^2 . Ці дисперсії обчислюють в генеральних сукупностях результатів вимірювань. У [5] встановлено, що середньому значенню μ випадкової величини, середньому вибіркочому \bar{x} необхідно приписувати відповідно такі середні квадратичні відхилення:

$$\sigma_{\mu} = \bar{\sigma} = \sqrt{2}\sigma; \tag{5}$$

$$s_{\bar{x}} = \bar{s} = \sqrt{2}s, \tag{6}$$

де $\bar{\sigma}$ називається середнім стандартним відхиленням, а \bar{s} – середнім вибіркочим стандартним відхиленням. Рівність (6) є наслідок з теореми (5).

Ціль статті: навести явні означення характеристик s^2, S^2 .

Виклад основного матеріалу досліджень. Припустимо, є сукупність результатів вимірювань $X = x_1, x_2, \dots, x_k$ обсягу k . Результати вимірювань будемо розглядати як множину елементів.

Впорядкованою парою елементів множини X називатиметься об'єкт $a_{ij} = (x_i, x_j)$, що складається з двох, не обов'язково різних елементів $x_i, x_j \in X$. Впорядкованими парами із X будуть також елементи $(x_i, x_i), (x_j, x_j)$, які будемо позначати так: $\Delta_{ii} = (x_i, x_i); \Delta_{jj} = (x_j, x_j)$.

Множина

$$C = \{(x_i, x_j) \mid x_i \in X, x_j \in X\}$$

впорядкованих пар елементів із X називається декартовим квадратом множини X і позначається через X^2 . Будь-яка підмножина впорядкованих пар елементів декартового квадрата називається бінарним відношенням на X . Декартовий квадрат має дві власні підмножини пар елементів: $a_{ij} = (x_i, x_j); a_{ji} = (x_j, x_i)$ і підмножину пар елементів в яких обидві координати рівні

$$\Delta = \{\Delta_{ii} \mid \Delta_{ii} = (x_i, x_i), x_i \in X\}. \tag{7}$$

Множина (7) називається відношенням рівності над множиною X , або діагоналлю, а елемент Δ_{ii} – нейтральним елементом [6].

Множина бінарних відношень на X відображена в табл. 1.

$C = X^2$ – декартовий квадрат вибірки

Множина бінарних відношень на x

$x_j \backslash x_i$	x_1	x_2	...	x_k
x_1	Δ_{11}	a_{12}	...	$a_{1.k}$
x_2	a_{21}	Δ_{22}	...	$a_{2.k}$
...
x_k	$a_{k.1}$	$a_{k.2}$...	$\Delta_{k.k}$

Множина різниць значень елементів бінарних відношень має обсяг $n' = k^2$ і виглядає так:

$$D' = \{d_{ij} = x_i - x_j; d_{ji} = x_j - x_i; \Delta'_{ii} = x_i - x_i \mid x_i \in X, x_j \in X\}$$

Особливість різниць впливає з такого прикладу.

П р и к л а д 1. Електронним тахеометром 3Та5Р чотири рази була виміряна одна коротка відстань. Одержано такий варіаційний ряд результатів вимірювань (мм):

$$X' = x'_1, x'_2, x'_3, x'_4 = 6832, 6834, 6835, 6835.$$

Значення ряду мають розмах $W = x'_4 - x'_1 = 3$ мм. Ступінь квантування значень $[Q] = x'_3 - x'_2 = 1$ мм. Від наведеного ряду перейдемо до такого ряду (мм):

$$X = x_1, x_2, x_3, x_4 = 2, 4, 5, 5. \quad (8)$$

Розглянемо три впорядковані пари елементів вибірки:

$$\Delta_{11} = (x_1, x_1) = (2, 2); a_{34} = (x_3, x_4) = (5, 5); a_{23} = (x_2, x_3) = (4, 5).$$

Знайдемо різниці значень елементів цих бінарних відношень (мм)

$$\Delta'_{11} = x_1 - x_1 = 2 - 2 = 0; \quad (9)$$

$$d_{34} = x_3 - x_4 = 5 - 5 = 0. \quad (10)$$

$$d_{23} = x_2 - x_3 = 4 - 5 = -1; \quad (11)$$

Сутність різниці (9), як і будь-яких різниць Δ'_{ii} така: з результатів вимірювань береться один результат вимірювання $x'_1 = 6832$ мм. Потім значення цього результату вимірювань порівнюється з самим собою

$$\Delta'_{11} = x'_1 - x'_1 = x_1 - x_1 = 6832 - 6832 = 0 \text{ мм.}$$

Отже, величина Δ'_{11} є нульова і являє собою нейтральну різницю. Проблема в тому, що всі нейтральні різниці значень елементів бінарних відношень, що утворюють діагональ декартового квадрата, враховуються при визначенні вибіркової дисперсії s^2 .

Величини (10), (11) – це різниці значень елементів вибірки.

Знайдемо суми квадратів від'ємних, додатних і нейтральних різниць значень елементів бінарних відношень декартового квадрата

$$\sum_{ij} d_{ij}^2 = \sum_{i=1}^{k-1} \sum_{j=i+1}^k (x_i - x_j)^2; \quad \sum_{ji} d_{ji}^2 = \sum_{i=1}^{k-1} \sum_{j=i+1}^k (x_j - x_i)^2;$$

$$\sum_{\Delta'_{ii}} \Delta_{ii}^2 = k \cdot 0 = 0.$$

Оскільки, $\sum_{ij} d_{ij}^2 = \sum_{ji} d_{ji}^2$, тому загальна сума квадратів різниць виглядає так:

$$\sum_{d'} d'^2 = 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^k (x_j - x_i)^2. \tag{12}$$

Знайдемо квадрат середньої квадратичної різниці значень елементів бінарних відношень на X

$$\overline{d'^2} = \frac{\sum_{d'} d'^2}{n'} = \frac{2 \sum_{i=1}^{k-1} \sum_{j=i+1}^k (x_j - x_i)^2}{k^2}. \tag{13}$$

Щоб визначити вибіркочну дисперсію s^2 , необхідно квадрат середньої квадратичної різниці (13) поділити на два.

Отже, приходимо до такого означення.

О з н а ч е н н я 1. *Вибіркова дисперсія s^2 являє собою половину квадрата середньої квадратичної різниці значень елементів бінарних відношень, що утворюють декартовий квадрат випадкової вибірки*

$$s^2 \equiv \frac{\overline{d'^2}}{2} = \frac{1}{k^2} \sum_{i=1}^{k-1} \sum_{j=i+1}^k (x_j - x_i)^2. \tag{14}$$

У табл. 2 наведено множину M бінарних відношень, що являє собою різницю декартового квадрата вибірки і діагоналі (7) цього квадрата

Таблиця 2

Множина $M = X^2 \setminus \Delta$

Множина M бінарних відношень

$x_i \backslash x_j$	x_1	x_2	...	x_k
x_1	ϕ	a_{12}	...	$a_{1.k}$
x_2	a_{21}	ϕ	...	$a_{2.k}$
...	ϕ	...
x_k	$a_{k.1}$	$a_{k.2}$...	ϕ

Множина M має обсяг $n = k(k-1)$. Її утворюють упорядковані пари значень результатів вимірювань, що подані в вибірці. Множина різниць вибірки має вигляд

$$D = \{d_{ij} = x_i - x_j; d_{ji} = x_j - x_i \mid x_i \in X, x_j \in X\}$$

Сума квадратів різниць, що є елементами множини D , дорівнює сумі (12). Знайдемо квадрат середньої квадратичної різниці значень вибірки

$$\overline{d^2} = \frac{\sum_d d^2}{n} = \frac{\sum_{d'} d'^2}{n} = \frac{2 \sum_{i=1}^{k-1} \sum_{j=i+1}^k (x_j - x_i)^2}{k(k-1)}.$$

Щоб визначити вибірку дисперсію S^2 необхідно квадрат середньої квадратичної різниці поділити на два. Отже, приходимо до такого означення.

О з н а ч е н н я 2. Є випадкова вибірка $X = x_1, x_2, \dots, x_k$ обсягу k . Вибірковою дисперсією S^2 називається половина квадрата середньої квадратичної різниці значень вибірки

$$S^2 \equiv \frac{\overline{d^2}}{2} = \frac{1}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^k (x_j - x_i)^2. \quad (15)$$

П р и к л а д 2. Візьмемо варіаційний ряд (8)

$$X = x_1, x_2, x_3, x_4 = 2, 4, 5, 5.$$

Щоб визначити вибірку дисперсії, обчислимо вибіркоче середнє значення (мм)

$$\bar{x} = \frac{1}{k} \sum_{i=1}^k x_i = \frac{2+4+5+5}{4} = \frac{16}{4} = 4,$$

квадрати відхилень значень ряду від вибіркового середнього \bar{x} (мм²)

$$(x_1 - \bar{x})^2 = (2 - 4)^2 = (-2)^2 = 4; (x_2 - \bar{x})^2 = (4 - 4)^2 = (0)^2 = 0;$$

$$(x_3 - \bar{x})^2 = (5 - 4)^2 = (1)^2 = 1; (x_4 - \bar{x})^2 = (5 - 4)^2 = (1)^2 = 1.$$

Знайдемо суму квадратів відхилень

$$\sum_{i=1}^k (x_i - \bar{x})^2 = 4 + 0 + 1 + 1 = 6 \text{ мм}^2.$$

За неявними означеннями (1), (2) обчислимо вибірку дисперсії (мм²)

$$s^2 = \frac{1}{k} \sum_{i=1}^k (x_i - \bar{x})^2 = \frac{1}{4} 6 = \frac{3}{2} = 1,5;$$

$$S^2 = \frac{1}{k-1} \sum_{i=1}^k (x_i - \bar{x})^2 = \frac{1}{4-1} 6 = \frac{6}{3} = 2.$$

Для повного обґрунтування отриманих наукових результатів визначимо ці ж дисперсії відповідно за різницями значень елементів бінарних відношень декартового квадрата вибірки і різницями значень результатів вимірювань.

П р и к л а д 3. Розглянемо декартовий квадрат вибірки (мм)

$$C = X^2 = \begin{vmatrix} \Delta_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & \Delta_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & \Delta_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & \Delta_{44} \end{vmatrix} = \begin{vmatrix} (2,2) & (2,4) & (2,5) & (2,5) \\ (4,2) & (4,4) & (4,5) & (4,5) \\ (5,2) & (5,4) & (5,5) & (5,5) \\ (5,2) & (5,4) & (5,5) & (5,5) \end{vmatrix}.$$

Знайдемо різниці значень елементів бінарних відношень (мм)

$$D' = \begin{vmatrix} \Delta'_{11} & d_{12} & d_{13} & d_{14} \\ d_{21} & \Delta'_{22} & d_{23} & d_{24} \\ d_{31} & d_{32} & \Delta'_{33} & d_{34} \\ d_{41} & d_{42} & d_{43} & \Delta'_{44} \end{vmatrix} = \begin{vmatrix} (2-2) & (2-4) & (2-5) & (2-5) \\ (4-2) & (4-4) & (4-5) & (4-5) \\ (5-2) & (5-4) & (5-5) & (5-5) \\ (5-2) & (5-4) & (5-5) & (5-5) \end{vmatrix} = \begin{vmatrix} 0 & -2 & -3 & -3 \\ 2 & 0 & -1 & -1 \\ 3 & 1 & 0 & 0 \\ 3 & 1 & 0 & 0 \end{vmatrix}.$$

Обчислимо квадрати різниць, їх суму (мм²)

$$|d'^2| = \begin{vmatrix} \Delta'^2_{11} & d'^2_{12} & d'^2_{13} & d'^2_{14} \\ d'^2_{21} & \Delta'^2_{22} & d'^2_{23} & d'^2_{24} \\ d'^2_{31} & d'^2_{32} & \Delta'^2_{33} & d'^2_{34} \\ d'^2_{41} & d'^2_{42} & d'^2_{43} & \Delta'^2_{44} \end{vmatrix} = \begin{vmatrix} 0 & 4 & 9 & 9 \\ 4 & 0 & 1 & 1 \\ 9 & 1 & 0 & 0 \\ 9 & 1 & 0 & 0 \end{vmatrix}.$$

$$\sum_{d'} d'^2 = 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^k (x_j - x_i)^2 = (0)(6) + (4)(2) + (9)(4) + (1)(4) = 0 + 8 + 36 + 4 = 48. \quad (16)$$

При обчисленні квадрата середньої квадратичної різниці значень елементів бінарних відношень враховують загальний обсяг $n' = k^2$ квадратів різниць значень результатів вимірювань і квадратів нейтральних різниць

$$\overline{d'^2} = \frac{\sum_{d'} d'^2}{n'} = \frac{\sum_{d'} d'^2}{k^2} = \frac{48}{16} = 3 \text{ мм}^2.$$

Щоб знайти вибіркочну дисперсію s^2 , поділимо квадрат середньої квадратичної різниці на два

$$s^2 \equiv \frac{\overline{d'^2}}{2} = \frac{3}{2} = 1,5 \text{ мм}^2;$$

$$s = \sqrt{s^2} = \sqrt{1,5} = 1,2 \text{ мм}.$$

За твердженням (6) обчислимо відхилення вибіркового середнього \bar{x}

$$s_{\bar{x}} = \bar{s} = \sqrt{2}s = \sqrt{2}\sqrt{1,5} = \sqrt{3} = 1,7 \text{ мм}.$$

Якщо є генеральна сукупність результатів вимірювань, то середнє значення μ визначається за означенням (4). Оскільки μ не є дійсною величиною, тому йому приписується відхилення (5). У наведеній випадковій вибірці положення очікуваного середнього значення μ оцінюється інтервалом (мм)

$$\bar{x} - s_{\bar{x}} \leq \mu \leq \bar{x} + s_{\bar{x}} = (4 - 1,7) \leq \mu \leq (4 + 1,7) = 2,3 \leq \mu \leq 5,7.$$

Розглянемо множину M бінарних відношень, що являє собою різницю декартового квадрата вибірки і діагоналі (7)

$$M = X^2 \setminus \Delta = \begin{vmatrix} \phi & a_{12} & a_{13} & a_{14} \\ a_{21} & \phi & a_{23} & a_{24} \\ a_{31} & a_{32} & \phi & a_{34} \\ a_{41} & a_{42} & a_{43} & \phi \end{vmatrix} = \begin{vmatrix} \phi & (2,4) & (2,5) & (2,5) \\ (4,2) & \phi & (4,5) & (4,5) \\ (5,2) & (5,4) & \phi & (5,5) \\ (5,2) & (5,4) & (5,5) & \phi \end{vmatrix}.$$

Сума квадратів різниць значень вибірки дорівнює сумі (16), тобто $\sum_d d^2 = \sum_{d'} d'^2 = 48$ мм².

Під час обчислення вибіркової дисперсії S^2 враховують обсяг $n = k(k-1)$ квадратів різниць значень вибірки

$$\overline{d^2} = \frac{\sum_d d^2}{n} = \frac{\sum_d d^2}{k(k-1)} = \frac{48}{4(4-1)} = \frac{48}{12} = 4 \text{ мм}^2;$$

$$S^2 \equiv \frac{\overline{d^2}}{2} = \frac{4}{2} = 2 \text{ мм}^2;$$

$$S = \sqrt{S^2} = \sqrt{2} = 1,4 \text{ мм.}$$

Визначимо середню квадратичну різницю значень результатів вимірювань

$$d = \sqrt{\overline{d^2}} = \sqrt{2}S = \sqrt{2}\sqrt{2} = 2 \text{ мм.}$$

Висновки

1. Правила (14), (15) – це явні означення двох вибірових дисперсій.
2. Характеристики \bar{x}, s^2, S^2 результатів вимірювань обчислюють в невеликих за обсягом сукупностях, а характеристики μ, σ^2, Σ^2 – в генеральних сукупностях, тобто в сукупностях, що мають повну групу G значень, наповненість $F = 1$, таку добротність Q , що дозволяє надійно встановити функцію $f(x)$ розподілу ймовірностей [7].

3. Дисперсія s^2 є оцінка дисперсії σ^2 . Вибіркове стандартне відхилення $s = \sqrt{s^2}$ не можна розглядати як оцінку розсіювання вибіркового середнього \bar{x} , оскільки, згідно з (2), воно є середнім квадратичним відхиленням значень $X = x$ вибірки від вибіркового середнього \bar{x} . Це записується так:

$$\bar{x} - s \leq X \leq \bar{x} + s.$$

4. Середнє вибірконе стандартне відхилення $\bar{s} = \sqrt{2}s$, що визначається за правилом (6) – це характеристика розсіювання можливих значень величини \bar{x} , являє собою середнє квадратичне відхилення цієї величини, приписується вибіркому середньому \bar{x} . За величиною \bar{s} визначають інтервал розсіювання очікуваного середнього значення μ

$$\bar{x} - \bar{s} \leq \mu \leq \bar{x} + \bar{s}.$$

5. Дисперсія S^2 – це оцінка дисперсії Σ^2 . Характеристика $S = \sqrt{S^2}$ не є середнім квадратичним відхиленням значень $X = x$ вибірки від вибіркового середнього \bar{x} , оскільки в (1) суму k квадратів відхилень $x_i - \bar{x}$ поділено не на k , а на $k - 1$.

Згідно з явним означенням (15), за величиною S оцінюються не відхилення результатів вимірювань, а їх різниці.

6. Точність вимірювань краще оцінювати не за відхиленнями, а за різницями значень результатів вимірювань. В [7] середню квадратичну різницю $d = \sqrt{2}S = T$ запропоновано називати оцінкою T для точності $T_0 = \sqrt{2}\Sigma$ вимірювань.

7. Підвищити точність вимірювань можна зменшенням розмаху $W = x_{\max} - x_{\min}$ і ступенем квантування $[Q]$ значень результатів вимірювань.

Перспективи подальших розвідок у цьому напрямку полягають в наведенні явних означень таких характеристик випадкових величин, які теорія точності вимірювань означає неявно.

1. *Философский энциклопедический словарь / Редкол.: С.С. Аверинцев, Э.А. Араб-Оглы, Л.Ф. Ильичев и др. – 2-е изд. – М.: Сов. энциклопедия, 1989. – С. 445.* 2. *Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. – М.: Наука, 1968. – 720 с.* 3. *Walpole Ronald E, Myers Raymond H. Probability and Statistics for Engineers and Scientists. 3-th edition, Macmillan Publishing Company. – New York, 1985. – 639 p.* 4. *Пряха Б.Г. Явні означення дисперсій // Новітні досягнення геодезії, геоінформатики та землевпорядкування – Європейський досвід. – Чернігів: КП “Видавництво ”Чернігівські обереги”, 2007. – С. 59–68.* 5. *Пряха Б. Оцінювання середніх значень // Сучасні досягнення геодезичної науки та виробництва, 2007 (I випуск): Зб. наук. праць. – Львів– С. 140–145.* 6. *Калужнин Л.А. Введение в общую алгебру. – М.: Наука, 1973. – 448 с.* 7. *Пряха Б.Г. Білецький Я.В. Про точність геодезичних вимірювань // Вісник геодезії та картографії. – 2003. – №3. – С. 43–49.*