Лекція 25. Методи знаходження точкових оцінок

План

1. Методи визначення точкових статистичних оцінок

2. Властивості $\bar{x}_B, D_{\text{В}}$

1. Методи визначення точкових статистичних оцінок

Існують три методи визначення точкових статистичних оцінок для параметрів генеральної сукупності.

Метод аналогій. Цей метод базується на тому, що для параметрів генеральної сукупності вибирають такі самі параметри вибірки, тобто для оцінки $\bar{X}_{\Gamma} = M\left(X\right), D_{\Gamma}$ вибирають аналогічні статистики — $\bar{x}_{\text{В}}, D_{\text{В}}$.

Метод найменших квадратів. Згідно з цим методом статистичні оцінки визначаються з умови мінімізації суми квадратів відхилень варіант вибірки від статистичної оцінки θ^*.

Отож, використовуючи метод найменших квадратів, можна, наприклад, визначити статистичну оцінку для $\bar{X}_{\Gamma} = M\left(X\right)$. Для цього скористаємося функцією $u = \sum_{i=0}^{n}\left(x_i - \theta^*\right)^2 n_i$. Використовуючи умову екстремуму, дістанемо:

$$\frac{\partial u}{\partial \theta^*} = -2\sum_{i=0}^{n}\left(x_i - \theta^*\right) n_i = 0 \rightarrow$$

$$\rightarrow \sum_{i=1}^{n} x_i n_i - \sum_{i=1}^{n} n_i \theta^* = 0 \rightarrow \theta^* = \frac{\sum_{i=1}^{n} x_i n_i}{n} = \bar{x}_{\text{В}}.$$

Звідси для $\theta = \bar{X}_{\Gamma}$ точковою статистичною оцінкою буде $\theta^* = \bar{x}_{\text{В}}$ — вибіркова середня.

Метод максимальної правдоподібності. Цей метод посідає центральне місце в теорії

Нехай ознака генеральної сукупності X визначається лише одним параметром θ і має щільність імовірностей $f(x;\ \theta)$. У разі реалізації вибірки з варіантами $x_1, x_2, ..., x_n$ щільність імовірностей вибірки буде такою:

$$f\left(x_1, x_2, ..., x_n, \theta^*\right) = f\left(x_1, \theta^*\right) \cdot f\left(x_2, \theta^*\right) \cdot ... \cdot f\left(x_n, \theta^*\right). \qquad (25.1)$$

При цьому варіанти розглядаються як незалежні випадкові величини, котрі мають один і той самий закон розподілу, що й ознака генеральної сукупності X.

Суть цього методу полягає в тому, що, фіксуючи значення варіант $x_1, x_2, ..., x_n$, визначають таке значення параметра θ^*, при якому функція (25.1) максимізується. Вона називається *функцією максимальної правдоподібності* і позначається так: $L = L\left(\theta^*\right)$.

Наприклад, коли ознака генеральної сукупності X має нормальний закон розподілу, то функція максимальної правдоподібності набере такого вигляду:

$$f\left(x_1, x_2, ..., x_n, \theta_1^*, \theta_2^*\right) = \frac{1}{\left(2\pi\theta_2\right)^{\frac{n}{2}}} \cdot e^{-\frac{\sum_{i=1}^{n}\left(x_i - \theta_1^*\right)^2}{2\theta_2^*}} \quad (25.2)$$

При цьому за статистичні оцінки θ_1^*, θ_2^* вибирають ті їх значення, за яких задана вибірка буде найімовірнішою, тобто функція (25.2) досягає максимуму. На практиці зручно від функції (25.2) перейти до її логарифма, а саме:

$$\ln f(x_1, x_2, ..., x_n, \theta_1^*, \theta_2^*) =$$

$$= L\left(x_1, x_2, ..., x_n, \theta_1^*, \theta_2^*\right) = -\frac{n}{2}\left(\ln\pi + \ln\theta_2^*\right) - \frac{\sum\left(x_i - \theta_1^*\right)^2}{2\theta_2^*}.$$

Згідно з необхідною умовою екстремуму для цієї функції дістанемо:

$$\begin{cases} \dfrac{\partial L}{\partial\theta_1^*} = -\dfrac{1}{\theta_2^*}\sum_{i=1}^{n}\left(x_i - \theta_1^*\right) = 0, \\ \dfrac{\partial L}{\partial\theta_2^*} = -\dfrac{n}{2\theta_2^*} + \dfrac{1}{2\left(\theta_2^*\right)^2} \cdot \sum_{i=1}^{n}\left(x_i - \theta_1^*\right)^2 = 0. \end{cases} \quad (25.3)$$

З першого рівняння системи (25.3) дістанемо:

$$\theta_1^* = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \overline{x}_{\text{В}}; \quad (25.4)$$

з другого рівняння системи (25.3) маємо:

$$\theta_2^* = \frac{1}{n} \cdot \sum_{i=1}^{n}\left(x_i - \overline{x}_{\text{В}}\right)^2 = D_{\text{В}}. \quad (25.5)$$

Отже, для $\overline{X}_{\Gamma} = M\left(X\right)$ точковою статистичною оцінкою є $\overline{x}_{\text{В}}$ для $D_{\Gamma} - D_{\text{В}}$.

2. Властивості $\overline{x}_B, D_{\text{В}}$

Виправлена дисперсія, виправлене середнє квадратичне відхилення. Точковою незміщеною статистичною оцінкою для $\overline{X}_{\Gamma} = M\left(X\right)$ є $\overline{x}_{\text{В}}$.

Отже, $M\left(\overline{x}_{\text{В}}\right) = \overline{X}_{\Gamma}$.

$D_{\text{В}}$ є точковою зміщеною статистичною оцінкою для D_{Γ}, де $\frac{n-1}{n}$ — коефіцієнт зміщення, який зменшується зі збільшенням обсягу вибірки *n*.

Отже, $\frac{n}{n-1} D_{\text{В}}$ буде точковою незміщеною статистичною оцінкою для D_{Γ}. Її назвали *виправленою дисперсією* і позначили через S^2.

Звідси точковою незміщеною статистичною оцінкою для D_{Γ} є виправлена дисперсія $S^2 = \frac{n}{n-1} D_{\text{В}}$ або

$$S^2 = \frac{n}{n-1} \cdot \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}_{\text{В}}\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}_{\text{В}}\right)^2}{n-1}. \quad (25.6)$$

Величину

$$S = \sqrt{\frac{n}{n-1} D_{\text{В}}} \qquad (25.7)$$

називають *виправленим середнім квадратичним відхиленням*.

Приклад. 200 однотипних деталей були піддані шліфуванню. Результати вимірювання наведені як дискретний статистичний розподіл, поданий у табличній формі:

x_i, мм	3,7	3,8	3,9	4,0	4,1	4,2	4,3	4,4
n_i	1	22	40	79	27	26	4	1

Знайти точкові незміщені статистичні оцінки для $\bar{X}_{\Gamma} = M(x)$, D_{Γ}.

Розв'язання. Оскільки точковою незміщеною оцінкою для $\bar{X}_{\Gamma}$ є $\bar{x}_{\text{В}}$, то обчислимо

$$\bar{x}_{\text{В}} = \frac{\sum x_i n_i}{n} =$$

$$= \frac{3,7 \cdot 1 + 3,8 \cdot 22 + 3,9 \cdot 40 + 4,0 \cdot 79 + 4,1 \cdot 27 + 4,2 \cdot 26 + 4,3 \cdot 4 + 4,4 \cdot 1}{200} =$$

$$= \frac{3,7 + 83,6 + 156 + 316 + 110,7 + 109,2 + 17,2 + 4,4}{200} = \frac{808,8}{200} = 4,004 \text{ мм.}$$

Для визначення точкової незміщеної статистичної оцінки для D_{Γ} обчислимо $D_{\text{В}}$:

$$\frac{\sum x_i^2 n_i}{n} = \frac{(3,7)^2 \cdot 1 + (3,8)^2 \cdot 22 + (3,9)^2 \cdot 40 + (4,0)^2 \cdot 79 +}{200} =$$

$$\frac{+(4,1)^2 \cdot 27 + (4,2)^2 \cdot 26 + (4,3)^2 \cdot 4 + (4,4)^2 \cdot 1}{200} =$$

$$= \frac{13,69 + 317,68 + 608,4 + 1264 + 453,87 + 458,64 + 73,96 + 19,36}{200} =$$

$$= \frac{3209,6}{200} = 16,048.$$

$$D_{\text{В}} = \frac{\sum x_i^2 n_i}{n} - (\bar{x}_{\text{В}})^2 = 16,048 - (4,004)^2 =$$

$$= 16,048 - 16,032016 = 0,015984.$$

Тоді точкова незміщена статистична оцінка для D_{Γ} дорівнюватиме:

$$S^2 = \frac{n}{n-1} D_{\text{В}} = \frac{200}{200-1} \cdot 0,015984 = \frac{200}{199} \cdot 0,015984 = 0,01606 \text{ мм}^2.$$