Лекція 6. Інтервальне оцінювання параметрів генеральної сукупності

При заміні істинного значення параметра розподілу генеральної сукупності θ його точковою оцінкою θ^* потрібно знати можливу похибку, яка виникає при використанні такої оцінки.

Для того, щоб мати уявлення про точність і надійність оцінки θ^* параметра θ використовують інтервальні статистичні оцінки.

При інтервальному оцінюванні вказують такий інтервал, що покриває оцінюваний параметр θ генеральної сукупності з заданою ймовірністю γ. Величину γ вибирають заздалегідь.

Нехай для заданого $\gamma \in (0;1)$ існує таке $\delta > 0$, що

$$P\left(\left|\theta^* - \theta\right| < \delta\right) = \gamma \quad \left(P\left(\theta^* - \delta < \theta < \theta^* + \delta\right) = \gamma\right)$$

Тоді випадковий інтервал $\left(\theta^* - \delta; \theta^* + \delta\right)$ називають *надійним* (або *довірчим*) інтервалом, число γ – *надійністю* (або *надійним рівнем*), δ - *точністю оцінки*, значення та $\theta^* - \delta$, $\theta^* + \delta$ – відповідно *нижньою і верхньою надійними межами*.

6.1. Надійні інтервали для параметрів нормального закону при відомому значенні σ_Γ із заданою надійністю γ

Нехай ознака *X* генеральної сукупності має нормальний закон розподілу $N(a;\sigma)$. Побудуємо довірчий інтервал для математичного сподівання, знаючи числове значення середнього квадратичного відхилення генеральної сукупності σ_Γ, із заданою надійністю γ. Оскільки випадкова величина $\bar{x}_B$, як точкова незміщена статистична оцінка для $\bar{X}_\Gamma = M(X)$, має нормальний закон розподілу $N\left(a; \dfrac{\sigma_\Gamma}{\sqrt{n}}\right)$, то дістанемо $P\left(\left|\bar{x}_B - a\right| < \delta\right) = \gamma$, де випадкова величина $\bar{x}_B - a$ має нормальний закон розподілу.

Випадкова величина $\dfrac{\overline{x}_B - a}{\dfrac{\sigma_\Gamma}{\sqrt{n}}}$ матиме стандартний нормальний $N(0;1)$.

Позначимо $\dfrac{\delta}{\dfrac{\sigma_\Gamma}{\sqrt{n}}} = x$ і перепишемо $P\left(\left|\theta^* - \theta\right| < \delta\right) = \gamma$ у вигляді

$$P\left(\left|\frac{\overline{x}_B - a}{\frac{\sigma_\Gamma}{\sqrt{n}}}\right| < x\right) = \gamma \tag{6.1}$$

або

$$P\left(\overline{x}_B - \frac{x \cdot \sigma_\Gamma}{\sqrt{n}} < a < \overline{x}_B + \frac{x \cdot \sigma_\Gamma}{\sqrt{n}}\right) = \gamma .$$

Згідно з формулою нормованого закону $P\left(\left|X - a\right| < \delta\right) = 2\Phi(\delta)$ для (6.1) вона набирає вигляду:

$$P\left(\left|\frac{\overline{x}_B - a}{\frac{\sigma_\Gamma}{\sqrt{n}}}\right| < x\right) = 2\Phi(x) - 1 = \gamma , \tag{6.2}$$

де $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left\{-\dfrac{z^2}{2}\right\} dz$ - функція розподілу стандартної нормальної випадкової величини.

З рівності (6.2) знаходимо аргументи x, а саме, за значенням функції Лапласа $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_{0}^{x} \exp\left\{-\dfrac{z^2}{2}\right\} dz$ (таблиця 2) використавши при цьому формулу:

$$\Phi(x) = 0,5 \cdot \gamma = \frac{\gamma}{2} \tag{6.3}$$

Шуканий довірчий інтервал матиме вигляд:

$$\overline{x}_B - \frac{x \cdot \sigma_\Gamma}{\sqrt{n}} < a < \overline{x}_B + \frac{x \cdot \sigma_\Gamma}{\sqrt{n}} \tag{6.4}$$

Величина $\dfrac{x\cdot\sigma_{\Gamma}}{\sqrt{n}}=\delta$ називається *точністю оцінки*, або *похибкою вибірки*.

Приклад 6.1. Знайти інтервальну оцінку з надійністю $\gamma=0,95$ для математичного сподівання a нормально розподіленої генеральної сукупності, якщо відомо $\sigma_{\Gamma}=5$, а за вибіркою об'єму $n=25$ знайшли вибіркове середнє $\bar{x}_B=14$.

Розв'язання: За умовою задачі маємо: $\bar{x}_B=14$, $\sigma_{\Gamma}=5$, $n=25$. Величину x знаходимо за табл. 2 як корінь рівняння $\Phi(x)=\dfrac{\gamma}{2}$, тобто $\Phi(x)=0,475$. Маємо $x=1,96$.

Знайдемо числові значення нижньої і верхньої надійних меж:

$$\bar{x}_B-\frac{x\cdot\sigma_{\Gamma}}{\sqrt{n}}=14-\frac{1,96\cdot 5}{\sqrt{25}}=12,04.$$

$$\bar{x}_B+\frac{x\cdot\sigma_{\Gamma}}{\sqrt{n}}=14+\frac{1,96\cdot 5}{\sqrt{25}}=15,96.$$

Підставимо у формулу (6.4): $12,04<a<15,96$.

Отже, з надійністю 0,95 (95% гарантії) математичне сподівання a нормально розподіленої генеральної сукупності покривається інтервалом $(12,04;15,96)$.

Приклад 6.2. Знайти мінімальний об'єм вибірки при якому з надійністю 0,975 точність оцінки математичного сподівання a нормально розподіленої генеральної сукупності $\delta=0,3$, якщо відомо $\sigma_{\Gamma}=1,2$.

Розв'язання. Похибка вибірки знаходиться за формулою

$$\delta=\frac{x\cdot\sigma_{\Gamma}}{\sqrt{n}}, \text{ звідси } n=\frac{x^2\cdot\sigma_{\Gamma}^2}{\delta^2}.$$

Знайдемо корінь рівняння $\Phi(x)=\frac{\gamma}{2}$, де $\gamma=0{,}975$. Звідси $\Phi(x)=\frac{0{,}975}{2}=0{,}4875$. Згідно табл. 2 $x=2{,}24$.

Обчислюємо мінімальний об'єм вибірки

$$n=\frac{2{,}24^2\cdot 1{,}2^2}{0{,}3^2}=81.$$

Зауваження. При малих вибірках $(n<30)$ для оцінювання математичного сподівання якщо невідоме значення середнього квадратичного відхилення, застосовується випадкова величина t, що має розподіл Стьюдента.

6.2. Надійні інтервали для параметрів нормального закону при невідомому значенні $\sigma_Г$ із заданою надійністю γ

Для побудови довірчого інтервалу, який оцінює математичне сподівання a нормально розподіленої генеральної сукупності при невідомому середньоквадратичному відхиленні із заданою надійністю γ, застосовується випадкова величина

$$t=\frac{\overline{x}_B-a}{\frac{S}{\sqrt{n}}}, \qquad (6.5)$$

що має розподіл Стьюдента з $k=n-1$ ступенями свободи.

Спочатку обчислюємо за даним статистичним розподілом вибіркове середнє $\overline{x}_B$, виправлене середнє квадратичне відхилення S, об'єм вибірки n і визначаємо за таблицею розподілу Стьюдента для оцінки математичного сподівання (табл. 3) значення $t_\gamma=t(\gamma,n)$ при заданій надійності (γ) та об'єму вибірки $(k=n)$, або значення $t_\gamma=t(\alpha,k)$ при заданому рівню значущості $(\alpha=1-\gamma)$ та числа ступенів свободи $(k=n-1)$ згідно табл. 6.

Шуканий надійний інтервал має вигляд:

$$\left(\bar{x}_B - \frac{t_\gamma \cdot S}{\sqrt{n}} < a < \bar{x}_B + \frac{t_\gamma \cdot S}{\sqrt{n}}\right) \quad (6.6)$$

Величина $\frac{t_\gamma \cdot S}{\sqrt{n}} = \delta$ називається *точністю оцінки*, або *похибкою вибірки*.

Приклад 6.3. З надійністю $\gamma = 0{,}95$ побудувати довірчий інтервал вигляду (6.6) за вибіркою 4; 2; 1; -2; 3; -2; 2; 5; 4; 3.

Розв'язання. Знайдемо вибіркове середнє і виправлене середнє квадратичне відхилення. Для спрощення обчислень побудуємо дискретний статистичний розподіл вибірки

x_i	-2	1	2	3	4	5
n_i	2	1	2	2	2	1

Обчислюємо:

$$\bar{x}_B = \frac{\sum_{i=1}^{6} x_i n_i}{n} = \frac{-2 \cdot 2 + 1 \cdot 1 + 2 \cdot 2 + 3 \cdot 2 + 4 \cdot 2 + 5 \cdot 1}{10} = \frac{20}{10} = 2.$$

$$D_B = \frac{\sum_{i=1}^{6} x_i^2 n_i}{n} - \bar{x}_B^2 = \frac{-2^2 \cdot 2 + 1^2 \cdot 1 + 2^2 \cdot 2 + 3^2 \cdot 2 + 4^2 \cdot 2 + 5^2 \cdot 1}{10} - 2^2 = 5{,}2.$$

$$S = \sqrt{\frac{n}{n-1} D_B} = \sqrt{\frac{10}{10-1} \cdot 5{,}2} \approx 2{,}4.$$

За таблицею значень розподілу Стьюдента за табл. 6, де $\alpha = 1 - 0{,}95 = 0{,}05$ та $k = n - 1 = 10 - 1 = 9)$ знаходимо значення $t(\alpha = 0{,}05; k = 9) = 2{,}26$. (Це ж саме значення можна отримати згідно табл. 3, де $t_\gamma(\gamma = 0{,}95; n = 10) = 2{,}26$)

Знайдемо числові значення нижньої і верхньої надійних меж:

$$\bar{x}_B - \frac{t_\gamma \cdot S}{\sqrt{n}} = 2 - \frac{2{,}26 \cdot 2{,}4}{\sqrt{10}} \approx 0{,}3.$$

$$\bar{x}_B + \frac{t_\gamma \cdot S}{\sqrt{n}} = 2 + \frac{2{,}26 \cdot 2{,}4}{\sqrt{10}} \approx 3{,}7.$$

Отже, з надійність $\gamma = 0{,}95$, можна стверджувати, що $0{,}3 < a < 3{,}7$.

Зауважимо, що при великих обсягах вибірки $(n > 30)$ на підставі центральної граничної теореми розподіл Стьюдента наближається до нормального закону розподілу і значення t_γ знаходиться за таблицею значень функції Лапласа $\Phi(x)$ (додаток 2).

6.3. Побудова довірчого інтервалу $\bar{X}_\Gamma$ із заданою надійністю γ для дисперсії D_Γ та середньоквадратичного відхилення σ_Γ

Спосіб перший через $\chi^2_{(\alpha;k)}$ (табл. 5)

Якщо досліджувана ознака $\bar{X}_\Gamma$ має нормальний закон розподілу, то для побудови довірчого інтервалу із заданою надійністю γ для D_Γ, σ_Γ використаємо випадкову величину

$$\chi^2 = \frac{n-1}{\sigma_\Gamma^2} S^2, \tag{6.7}$$

що має розподіл χ^2 із $k = n-1$ ступенями свободи.

Тобто, довірчий інтервал для D_Γ можна записати у вигляді:

$$\frac{n-1}{\chi_2^2} S^2 < D_\Gamma < \frac{n-1}{\chi_1^2} S^2, \tag{6.8}$$

звідки довірчий інтервал для σ_Γ:

$$\frac{\sqrt{n-1}}{\chi_2} S < \sigma_\Gamma < \frac{\sqrt{n-1}}{\chi_1} S \tag{6.9}$$

де значення χ_1^2, χ_2^2 знаходимо (за табл. 5) ймовірності $P\left(\chi^2 > \chi^2_{\alpha;k}\right)$.

Додатні значення χ_1^2, χ_2^2 знаходимо за табл. 5 з використанням рівностей:

$$F\left(\chi_1^2(\gamma)\right) = \frac{1+\gamma}{2} \text{ або } F\left(\chi_1^2(\alpha)\right) = 1 - \frac{\alpha}{2}. \tag{6.10}$$

$$F\left(\chi_2^2(\gamma)\right) = \frac{1-\gamma}{2} \text{ або } F\left(\chi_2^2(\alpha)\right) = \frac{\alpha}{2} \tag{6.11}$$

Приклад 6.4. За вибіркою 4; 2; 1; -2; 3; -2; 2; 5; 4; 3 з нормально розподіленої генеральної сукупності побудувати довірчі інтервали з надійністю $\gamma = 0,9$ для дисперсії $D_Г$ та середнього квадратичного відхилення $\sigma_Г$.

Розв'язання. Для побудови довірчих інтервалів скористаємось формулами (6.8) і (6.9). Виправлене середнє квадратичне відхилення було знайдено у попередньому прикладі: $S \approx 2,4$. Виправлена дисперсія $S^2 = 2,4^2 \approx 5,8$.

Згідно формул (6.10) та (6.11) знайдемо значення для χ_1^2 і χ_2^2 відповідно. Кількість ступенів свободи $k = n - 1 = 9$. При надійності $\gamma = 0,9$, рівень значущості буде $\alpha = 1 - \gamma = 1 - 0,9 = 0,1$. Отже знаходимо ймовірності:

$$F\left(\chi_1^2(\alpha)\right) = 1 - \frac{\alpha}{2} = 1 - \frac{0,1}{2} = 0,95 \text{ та } F\left(\chi_2^2(\alpha)\right) = \frac{\alpha}{2} = \frac{0,1}{2} = 0,05.$$

Згідно табл. 5 для ймовірностей 0,95 і 0,05 при $k = 9$ знаходимо $\chi_1^2 = 3,325$ та $\chi_2^2 = 16,919$. Тоді довірчий інтервал з надійністю для дисперсії за (6.8) має вигляд:

$$\frac{9}{16,919} \cdot 5,8 < D_Г < \frac{9}{3,325} \cdot 5,8 \quad \Rightarrow \quad 3,1 < D_Г < 15,8$$

Довірчий інтервал з надійністю $\gamma = 0,9$ для $\sigma_Г$ (6.9):

$$\sqrt{3,1} < \sigma_Г < \sqrt{15,8} \quad \Rightarrow \quad 1,8 < \sigma_Г < 4,0.$$

Спосіб другий через $\chi = q_{(\gamma, n)}$ (табл. 4)

Довірчий інтервал для $\sigma_Г$ із заданою надійністю γ можна побудувати з використанням розподілу χ.

За заданою надійністю (γ) і обсягом вибірки (n) знаходимо за табл. 4 значення величини $q = q(\gamma, n)$.

Якщо $q > 1$ тоді довірчий інтервал для $\sigma_Г$ записується у вигляді:

$$0 < \sigma_Г < S \cdot (1+q). \quad (5.12)$$

Якщо $q<1$ тоді довірчий інтервал для $\sigma_Г$ записується у вигляді:

$$S \cdot (1-q) < \sigma_Г < S \cdot (1+q). \quad (5.13)$$

Приклад 6.5. Побудувати довірчі інтервали для $\sigma_Г$ з надійностями $\gamma_1 = 0{,}95$ та $\gamma_2 = 0{,}99$ за вибіркою об'єму $n=10$, якщо знайдено $S = 2{,}4$.

Розв'язання. Для побудови потрібних довірчих інтервалів знайдемо за додатком 5 значення $q_1 = q(\gamma_1, n)$ та $q_2 = q(\gamma_2, n)$.

При $\gamma_1 = 0{,}95$, маємо $q_1 = 0{,}65 < 1$, тому довірчий інтервал з надійністю $\gamma_1 = 0{,}95$ запишему у вигляді (4.12), а з надійністю $\gamma_2 = 0{,}99$, та $q_2 = 1{,}08 > 1$ - у вигляді (4.13). Визначаємо величини:

$$S(1-q_1) = 2{,}4 \cdot (1-0{,}65) = 2{,}4 \cdot 0{,}35 = 0{,}84;$$
$$S(1+q_1) = 2{,}4 \cdot (1+0{,}65) = 2{,}4 \cdot 1{,}65 = 3{,}96.$$

$$S(1+q_2) = 2{,}4 \cdot (1+1{,}08) = 2{,}4 \cdot 2{,}08 = 4{,}99.$$

Отже, довірчий інтервал для $\sigma_Г$ з надійністю $\gamma_1 = 0{,}95$ має вигляд: $0{,}84 < \sigma_Г < 3{,}96$, з надійністю $\gamma_2 = 0{,}99$: $0 < \sigma_Г < 4{,}99$.

6.4. Довірчий інтервал для оцінки ймовірності настання події

1. Для оцінки генеральної частки p нормально розподіленої кількості ознак $X_Г$ за вибірковою часткою $\omega = \frac{m}{n}$ (*при великому обсязі вибірки*, тобто при $n \ge 30$) та власне-випадковому *повторному відборі* матимемо формулу:

$$P\left(\omega - t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n}} < p < \omega + t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n}}\right) = 2 \cdot \Phi(t) = \gamma,$$

де t визначається за табл. 2 функції Лапласа із співвідношення $2 \cdot \Phi(t) = \gamma$; ω - вибіркова частка; n - об'єм вибірки; $\Delta = t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n}}$ - гранична похибка.

2. Для оцінки генеральної частки p нормально розподіленої кількості ознак $X_Г$ за вибірковою часткою $\omega = \frac{m}{n}$ (*при великому обсязі вибірки*, тобто при $n \geq 30$) та власне-випадковому *безповторному відборі* матимемо формулу:

$$P\left(\omega - t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n} \cdot \left(1 - \frac{n}{N}\right)} < p < \omega + t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n} \cdot \left(1 - \frac{n}{N}\right)}\right) = 2 \cdot \Phi(t) = \gamma,$$

де N - обсяг генеральної сукупності; $\Delta = t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n} \cdot \left(1 - \frac{n}{N}\right)}$ - гранична похибка.

3. Для оцінки генеральної частки p нормально розподіленої кількості ознак $X_Г$ за вибірковою часткою $\omega = \frac{m}{n}$ (*при малому обсязі вибірки*, тобто при $n < 30$) та власне-випадковому *повторному відборі* матимемо формулу:

$$P\left(\omega - t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n}} < p < \omega + t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n}}\right) = 2 \cdot S(t) = \gamma,$$

де t визначається за табл. 6 розподілу Стьюдента при рівні значущості $\alpha = 1 - \gamma$ та числом ступенів свободи $k = n - 1$.

4. Для оцінки генеральної частки p нормально розподіленої кількості ознак $X_Г$ за вибірковою часткою $\omega = \frac{m}{n}$ (*при малому обсязі вибірки*, тобто при $n < 30$) та власне-випадковому *безповторному відборі* матимемо формулу:

$$P\left(\omega - t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n} \cdot \left(1 - \frac{n}{N}\right)} < p < \omega + t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n} \cdot \left(1 - \frac{n}{N}\right)}\right) = 2 \cdot S(t) = \gamma.$$

Приклад 6.6. Знайти довірчий інтервал для оцінки ймовірності p настання деякої події з надійністю $\gamma = 0{,}95$, якщо у 80 спробах ця подія настала 20 разів.

Розв'язання. За умовою $\gamma = 0{,}95$, $n = 80 > 30$, $m = 20$.

Для оцінки генеральної частки p обераємо формулу при $n \geq 30$ та власне-випадковому *повторному відборі*:

$$P\left(\omega - t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n}} < p < \omega + t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n}}\right) = 2 \cdot \Phi(t) = \gamma,$$

Корінь рівняння $2\Phi(t) = \gamma$ знаходимо із табл. 2:

$$\Phi(t) = \frac{0,95}{2} = 0,475; \quad t = 1,96.$$

Обчислюємо відносну частоту настання події $\omega = \frac{m}{n} = \frac{1}{4} = 0,25$.

Знайдемо граничну похибку згідно формули:

$$\Delta = t \cdot \sqrt{\frac{\omega \cdot (1-\omega)}{n}} = 1,96 \cdot \sqrt{\frac{0,25 \cdot (1-0,25)}{80}} \approx 0,0949.$$

Тоді згідно формули $\omega - \Delta < p < \omega + \Delta$, отримаємо:

$$0,25 - 0,0949 < p < 0,25 + 0,0949$$
$$0,155 < p < 0,345.$$

6.5. Побудова довірчого інтервалу для математичного сподівання за допомогою нерівності Чебишова із заданою надійністю γ

Якщо немає впевненості, що досліджувана ознака генеральної сукупності має нормальний розподіл, тоді для побудови довірчого інтервалу для математичного сподівання a із заданою надійністю γ використовують нерівність Чебишова:

$$P\left(\left|\overline{x}_B - a\right| < \delta\right) \geq 1 - \frac{\sigma_\Gamma^2}{n\delta^2} = \gamma \qquad (6.14)$$

Останню рівність розв'яжемо відносно точності оцінки δ:

$$\delta = \frac{\sigma_\Gamma}{\sqrt{(1-\gamma)n}} \qquad (6.15)$$

Отже, шуканий довірчий інтервал має вигляд:

$$\bar{x}_B - \frac{\sigma_\Gamma}{\sqrt{(1-\gamma)n}} < a < \bar{x}_B + \frac{\sigma_\Gamma}{\sqrt{(1-\gamma)n}} \quad (6.16)$$

Якщо теоретичне середнє квадратичне відхилення невідоме, то його замінюють виправленим вибірковим середнім квадратичним відхиленням S і довірчий інтервал для математичного сподівання a із заданою надійністю γ шукають у вигляді

$$\bar{x}_B - \frac{S}{\sqrt{(1-\gamma)n}} < a < \bar{x}_B + \frac{S}{\sqrt{(1-\gamma)n}} \quad (6.17)$$

Приклад 6.7. З надійністю $\gamma = 0{,}95$ побудувати довірчий інтервал для математичного сподівання a за допомогою нерівності Чебишова, якщо за вибіркою об'єму $n = 10$ знайдено $\bar{x}_B = 2$ і $S = 2{,}4$.

Розв'язання. Скористаємось формулою (6.17). Обчислимо

$$\frac{S}{\sqrt{(1-\gamma)n}} = \frac{2{,}4}{\sqrt{(1-0{,}95)\cdot 10}} = \frac{2{,}4}{0{,}7} \approx 3{,}4 .$$

Шуканий довірчий інтервал має вигляд: $2 - 3{,}4 < a < 2 + 3{,}4 \Rightarrow -1{,}4 < a < 5{,}4$.

Отже, з надійністю $\gamma = 0{,}95$ математичне сподівання $a \in (-1{,}4; 5{,}4)$.