Лекція 3

Інтервальне оцінювання.

3.1. Довірча ймовірність і гранична похибка вибірки

Вище розглянута оцінка параметрів θ генеральної сукупності одним числом, тобто $\overline{x}_Г$ – числом $\overline{x}_В$, p – числом w, σ^2 – числом s^2 або $\tilde{s}^2$. Такі оцінки параметрів називаються **точковими.** Проте точкова оцінка θ_n є лише наближеним значенням невідомого параметра θ навіть в тому випадку, якщо вона незміщена, грунтовна і ефективна, і для вибірки малого об'єму може істотно відрізнятись від θ. Для того, щоб отримати уявлення про точність і надійність оцінки θ_n параметра θ, використовують **інтервальну оцінку параметра**.

Інтервальною оцінкою параметра θ називається числовий інтервал $\left(\theta_n^{(1)}, \theta_n^{(2)}\right)$, який із заданою ймовірністю γ накриває невідоме значення параметра θ. Межі інтервалу $\left(\theta_n^{(1)}, \theta_n^{(2)}\right)$ і його границі знаходяться за вибірковими даними, і тому є випадковими величинами на відміну від параметра θ, який оцінюється – величини невипадкової. Тому доцільніше говорити про те, що інтервал $\left(\theta_n^{(1)}, \theta_n^{(2)}\right)$ «накриває», а не «містить» значення θ.

Інтервал $\left(\theta_n^{(1)}, \theta_n^{(2)}\right)$ називається **довірчим**, а ймовірність γ – **довірчою ймовірністю, рівнем довіри** або **надійністю оцінки.**

Величина довірчого інтервалу суттєво залежить від обсягу вибірки n (зменшується з ростом n) і від значення довірчої ймовірності γ (збільшується при наближенні γ до одиниці). Дуже часто (але не завжди) довірчий інтервал вибирається симетричним відносно параметра θ, тобто $\left(\theta-\Delta,\theta+\Delta\right)$.

Найбільше відхилення Δ оцінки θ_n від параметра, що оцінюється θ, зокрема, вибіркового середнього (або частки) від генерального середнього (або частки), яке можливе із заданою довірчою ймовірністю γ, називається **граничною похибкою вибірки.**

Похибка Δ є **похибкою репрезентативності** вибірки. Вона виникає тільки внаслідок того, що досліджується не вся сукупність, а лише її частина (вибірка), яка відібрана випадково.

3.2. Побудова довірчого інтервалу для генерального середнього і генеральної частки за великими вибірками

Теорема 3.1. Ймовірність того, що відхилення вибіркового середнього (чи частки) не перевищить число $\Delta>0$ (за абсолютною величиною), дорівнює:

$$P\left(\left|\bar{x}-\bar{x}_0\right|\le\Delta\right)=\Phi\left(t\right)=\gamma, \quad (3.1)$$

де $t=\dfrac{\Delta}{\sigma_x}$,

$$P\left(\left|w-p\right|\le\Delta\right)=\Phi(t)=\gamma, \quad (3.2)$$

де $t=\dfrac{\Delta}{\sigma_w}$,

$\Phi(t)$ - функція (інтеграл ймовірностей) Лапласа.

Доведення. Вибіркове середнє $\bar{x}$ і вибіркова частка w повторної вибірки є сумою n незалежних випадкових величин $\dfrac{\sum_{k=1}^{n}X_k}{n}=\sum_{k=1}^{n}\dfrac{X_k}{n}$, де $X_k\,(k=1,2,\dots,n)$ має один і той самий закон розподілу із скінченним математичним сподіванням і дисперсією. Звідси, на основі теореми Ляпунова при $n\to\infty$ розподіли $\bar{x}$ і w необмежено наближаються до нормальних (практично при $n>40$ розподіли $\bar{x}$ і w можна вважати наближено нормальними). Для безповторної вибірки $\bar{x}$ і w є сумою залежних випадкових величин. Можна показати, що і в цьому випадку при $n \to \infty$ закон розподілу $\overline{x}$ і w наближається до нормального.

Формули (3.1) і (3.2) випливають безпосередньо із властивості нормального закону. Ці формули отримали назву **формул довірчої ймовірності** для середнього та частки. **Доведено**.

Середнє квадратичне відхилення вибіркового середнього σ_x та вибіркової частки σ_w власно-випадкової вибірки називається **середньою квадратичною (стандартною) похибкою вибірки.**

Наслідок 1. При заданій довірчій ймовірності γ гранична похибка вибірки дорівнює *t*-кратній величині середньої квадратичної похибки, де $\Phi(t)=\gamma$, тобто

$$\Delta = t\sigma_{\bar{x}}, \tag{3.3}$$

$$\Delta = t\sigma_w . \tag{3.4}$$

Наслідок 2. Інтервальні оцінки (довірчі інтервали) для генерального середнього і генеральної частки можуть бути знайдені за формулами:

$$\overline{x} - \Delta \le \overline{x}_0 \le \overline{x} + \Delta, \tag{3.5}$$

$$w - \Delta \le p \le w + \Delta. \tag{3.6}$$

Оскільки генеральні частка p і дисперсія σ^2 невідомі, то в формулах табл. 3.1 заміняємо їх грунтовними оцінками за вибіркою – відповідно w і s^2, бо при достатньо великому обсязі вибірки n практично достовірно, що $w \approx p, s^2 \approx \sigma^2$.

Таблиця 3.1. Середні квадратичні похибки вибірки для частки та середнього

Оцінюваний параметр	Формули середніх квадратичних похибок вибірки	
	повторна вибірка	безповторна вибірка
Середнє	$\sigma_{\bar{x}} = \sqrt{\dfrac{\sigma^2}{n}} \approx \sqrt{\dfrac{s^2}{n}}$ (3.7)	$\sigma'_{\bar{x}} \approx \sqrt{\dfrac{s^2}{n}\left(1-\dfrac{n}{N}\right)}$ (3.8)
Частка	$\sigma_w = \sqrt{\dfrac{pq}{n}} \approx \sqrt{\dfrac{w(1-w)}{n}}$ (3.9)	$\sigma'_w \approx \sqrt{\dfrac{w(1-w)}{n}\left(1-\dfrac{n}{N}\right)}$ (3.10)

При визначенні середньої квадратичної похибки вибірки для частки, якщо навіть w невідома, в якості pq можна взяти його максимально можливе значення $(pq)_{max} = \left[p(1-p)\right]_{max} = 0{,}5 \cdot 0{,}5 = 0{,}25.$

Приклад 3.1. Тестується нова програма для розпізнавання прихованих зображень. Було протестовано 100 картинок із прихованим зображенням. Одержали наступні данні щодо часу дешифрування (у хвилинах) для кожного зі 100 прихованих зображень. В таблиці наведені ранжовані варіанти вибірки.

Необхідно визначити: а) ймовірність того, що середній час дешифрування зображення для заданої програми в цілому відрізняється від вибіркового середнього не більше, ніж на 1хвилину (за абсолютною величиною); б) межі, між якими з ймовірністю 0,9545 знаходиться середній час роботи програми.

Розв’язання. а) $n = 100$. Раніше в прикладі були обчислені $\bar{x} = 119{,}2(хв)$, $s^2 = 87{,}48$. Знайдемо середню квадратичну похибку вибірки для середнього:

$\sigma_{\bar{x}} = \sqrt{\dfrac{87{,}48}{100}} = 0{,}935(хв)$ (для повторної вибірки).

Тепер шукану довірчу ймовірність знаходимо за (3.1):

$$P\left(\left|\bar{x}-\bar{x}_0\right|\le 1\right)=\Phi\left(\frac{1}{0,935}\right)=\Phi(1,07)=0,715.$$

(Значення $\Phi(t)$ знаходимо за стандартною таблицею).

Отже, ймовірність того, що вибіркове середнє відрізняється від генерального середнього не більше, ніж на 1хв. (за абсолютною величиною), дорівнює 0,715 для повторної вибірки.

б) Знайдемо граничні похибки: (вибірка повторна) $\gamma=\Phi(t)=0,9545 \Rightarrow t=2,0$. $\Delta=2,00\cdot 0,935=1,870(хв)$

Шуканий довірчий інтервал визначаємо за (3.5):

$$119,2-1,870\le \bar{x}_0\le 119,2+1,870 \text{ або } 117,33\le \bar{x}_0\le 121,07\,(хв)$$

Таким чином, з надійністю 0,9545 середній час роботи програми, яка тестується, знаходиться в межах від 117,33 до 121,07 хвилин (вибірка повторна).►

3.3. Обсяг вибірки

Для проведення вибіркового спостереження досить важливо правильно визначити обсяг вибірки n, який значною мірою визначає необхідні при цьому часові, трудові і вартісні витрати. Для визначення n необхідно задати надійність (довірчу ймовірність) оцінки γ і точність (граничну похибку вибірки) Δ.

Обсяг вибірки знаходиться з формули, що виражає граничну похибку вибірки через дисперсію ознаки. Наприклад, для повторної вибірки формула має вигляд: $\Delta=t\sqrt{\frac{\sigma^2}{n}}$, звідки $n=\frac{t^2\sigma^2}{\Delta^2}$, де $\Phi(t)=\gamma$. Аналогічно можуть бути отримані й інші формули обсягу вибірки, які зведемо в таблицю 3.2.

Таблиця 3.2. Формули об'єму вибірки

Оцінюваний параметр	Повторна вибірка	Безповторна вибірка
Генеральне середнє	$n=\frac{t^2\sigma^2}{\Delta^2}$ (3.11)	$n'=\frac{Nt^2\sigma^2}{t^2\sigma^2+N\Delta^2}$ (3.12)
Генеральна частка	$n=\frac{t^2pq}{\Delta^2}$ (3.13)	$n'=\frac{Nt^2pq}{t^2pq+N\Delta^2}$ (3.14)

Для визначення обсягу вибірки необхідно знати характеристики генеральної сукупності σ^2 або p, які невідомі, і для визначення яких передбачаєтся проведення вибіркового спостереження. В якості цих характеристик зазвичай використовують вибіркові дані s^2 або w попереднього дослідження в аналогічних умовах, тобто вважають $\sigma^2 \approx s^2$(або $\tilde{s}^2$) або $p \approx w$.

При оцінці генеральної частки (якщо про неї нічого невідомо) замість проведення пробної вибірки можна в якості $pq = p(1-p)$ взяти його максимально можливе значення, рівне 0,25, але при цьому необхідно враховувати, що знайдене значення обсягу вибірки буде більшим від мінімально необхідного для заданих точності та надійності оцінок.

Приклад 3.2. За умовою прикладу 3.1 визначити обсяг вибірки, при якому із ймовірністю 0,9973 відхилення середнього часу роботи програми від генерального середнього не перевищить 1хвилину (за абсолютною величиною).

Розв'язання. В якості невідомого значення для визначення обсягу вибірки беремо її грунтовну оцінку $s^2 = 87{,}48$, знайдену раніше в прикладі. Враховуючи, що $\gamma = \Phi(t) = 0{,}9973$ і $t = 3{,}00$, знайдемо обсяг повторної вибірки за (3.11), тобто $n = 3^2 \cdot 87{,}48 / 1 = 787$. Отже, для заданої точності $\Delta = 1(\text{хв})$ і надійності $\gamma = 0{,}9973$ обсяг вибірки значно збільшився.►

3.4. Побудова довірчого інтервалу для генеральної дисперсії

Нехай розподіл ознаки (випадкової величини) X в генеральній сукупності є нормальним $N\left(\overline{x}_0;\sigma^2\right)$. Будемо думати, що математичне сподівання (генеральне середнє) відоме. Тоді вибіркова дисперсія повторної вибірки $X_1, X_2, ..., X_n$: $s_*^2 = \dfrac{\sum_{i=1}^{n}\left(X_i - \overline{x}_0\right)^2}{n}$ (її не варто плутати з вибірковою дисперсією $s^2 = \dfrac{\sum_{i=1}^{n}\left(X_i - \overline{x}_B\right)^2}{n}$ і «виправленою» вибірковою дисперсією $\tilde{s}^2 = \dfrac{n}{n-1}s^2$: якщо s_*^2 характеризує варіацію значень ознаки відносно генерального середнього $\overline{x}_0$, то s^2 і $\tilde{s}^2$ - відносно вибіркового середнього $\overline{x}$).

Розглянемо статистику $\chi^2 = \dfrac{ns_*^2}{\sigma^2} = \sum_{i=1}^{n} t_i^2$. Враховуючи, що $M\left(X_i\right) = \overline{x}_0$, $D\left(X_i\right) = \sigma^2$, $\left(i = 1, 2, ..., n\right)$, неважко показати, що $M\left(t\right) = 0$ і $\sigma_{t_i}^2 = D\left(t_i\right) = 1$.

Розподіл суми квадратів n незалежних випадкових величин $\sum_{i=1}^{n} t_i^2$, кожна з яких має стандартний нормальний розподіл $N\left(0;1\right)$, є розподілом χ^2 з $k = n$ ступенями свободи. Розподіл χ^2 не залежить від невідомих параметрів випадкової величини X, а залежить лише від числа ступенів свободи *k*.

Щільність ймовірностей розподілу χ^2 має складний вигляд, і її інтегрування є складним. Складено таблиці для обчислення ймовірності того, що випадкова величина (яка має χ^2- розподіл з k ступенями свободи) перевищить деякі критичні значення $\chi_{a;k}^2$, тобто $P\left(\chi^2 > \chi_{a;k}^2\right) = a$. В практиці вибіркового спостереження математичне сподівання $\overline{x}_0$, як правило, невідоме, і доводиться мати справу не з s_*^2 , а з s^2 . Якщо $X_1, X_2, ..., X_n$ - повторна вибірка із нормально розподіленої генеральної сукупності, то випадкова величина $\frac{ns^2}{\sigma^2}$ (або $\frac{(n-1)^2 \tilde{s}^2}{\sigma^2}$) має розподіл з $k = n-1$ ступенями свободи. Тому для заданої довірчої ймовірності γ можна записати: $P\left(\chi_1^2 < \frac{ns^2}{\sigma^2} < \chi_2^2\right) = \gamma$.

Очевидно, що значення χ_1^2 і χ_2^2 визначаються неоднозначно при одному і тому ж значенні γ. Зазвичай χ_1^2 і χ_2^2 вибирають таким чином, щоб ймовірності подій $\chi^2 < \chi_1^2$ і $\chi^2 > \chi_2^2$ були однаковими, тобто

$$P\left(\chi^2 < \chi_1^2\right) = P\left(\chi^2 > \chi_2^2\right) = \frac{1-\gamma}{2}.$$

Перетворивши подвійну нерівність $\chi_1^2 < \frac{ns^2}{\sigma^2} < \chi_2^2$ до рівносильного вигляду $\frac{ns^2}{\chi_2^2} < \sigma^2 < \frac{ns^2}{\chi_1^2}$, отримаємо формулу довірчої ймовірності для генеральної дисперсії: $P\left(\frac{ns^2}{\chi_2^2} < \sigma^2 < \frac{ns^2}{\chi_1^2}\right) = \gamma$, а для середнього квадратичного відхилення: $P\left(\frac{\sqrt{n}s}{\chi_2} < \sigma < \frac{\sqrt{n}s}{\chi_1}\right) = \gamma$.

При використанні таблиць значень $\chi_{a;k}^2$, отриманих з рівності $P\left(\chi^2 > \chi_1^2\right) = a$, необхідно врахувати , що $P\left(\chi^2 < \chi_1^2\right) = 1 - P\left(\chi^2 > \chi_1^2\right)$, тому умова $P\left(\chi^2 > \chi_1^2\right) = \frac{1+\gamma}{2}$ рівносильна умові $P\left(\chi^2 > \chi_1^2\right) = 1 - \frac{1-\gamma}{2} = \frac{1+\gamma}{2}$. Таким чином, значення χ_1^2 і χ_2^2 знаходимо по таблиці із рівностей: $P\left(\chi^2 > \chi_1^2\right) = \frac{1+\gamma}{2}$, $P\left(\chi^2 < \chi_2^2\right) = 1 - \frac{1-\gamma}{2}$.

Тобто при $k = n - 1$ $\quad \chi_1^2 = \chi^2_{(1+\gamma)/2;n-1}, \quad \chi_2^2 = \chi^2_{(1-\gamma)/2;n-1}$.

Приклад 3.3. Досліджуються статистичні характеристики великої комп'ютерної мережі. На основі вибіркових спостережень інтенсивності роботи 20 мережевих комп'ютерів було встановлено, що середнє квадратичне відхилення добового навантаження складає 15 Гб на годину. Вважаючи , що інтенсивність роботи комп'ютерів має нормальний розподіл, знайти межі, в яких з надійністю 0,9 знаходяться генеральна дисперсія і середнє квадратичне відхилення добового навантаження на один комп'ютер у всій мережі.

Розв'язання. Маємо $\gamma = 0{,}9; \left(1-\gamma\right)/2 = 0{,}05;\ \left(1+\gamma\right)/2 = 0{,}95$.

При числі ступенів свободи $k = n - 1$=20-1=19 визначимо χ_1^2 і χ_2^2 за таблицею значень для χ^2 - розподілу: $\chi_1^2 = \chi^2_{0{,}95;19} = 10{,}1$ і $\chi_2^2 = \chi^2_{0{,}05;19} = 30{,}1$.

Тоді довірчий інтервал для σ^2 можна записати у вигляді: $\frac{20}{30{,}1} \cdot 15^2 < \sigma^2 < \frac{20}{10{,}1} \cdot 15^2$ або $149{,}5 < \sigma^2 < 445{,}6$,

і для σ: $\sqrt{149{,}5} < \sigma < \sqrt{445{,}6}$ або $12{,}2 < \sigma < 21{,}1$ $\left(\text{Гб/год}\right)$.

Отже, з надійністю 0,9 дисперсія добового навантаження знаходиться в межах від 149,5 до 445,6 , а її середнє квадратичне відхилення – від 12, 2 до 21,1 Мб за годину.►

Зауваження. Таблиця значень $\chi^2_{a;k}$ складена для числа ступенів свободи k від 1 до 30. При $k > 30$ можна вважати, що випадкова величина $\sqrt{2\chi^2} - \sqrt{2k-1}$ має стандартний нормальний розподіл $N(0;1)$. Тому для визначення χ_1^2 і χ_2^2 необхідно записати, що $P\left(\left|\sqrt{2\chi^2} - \sqrt{2k-1}\right| < t\right) = \Phi(t) = \gamma$, звідки $-t < \sqrt{2\chi^2} - \sqrt{2k-1} < t$, і після перетворень:

$$\frac{1}{2}\left(\sqrt{2k-1} - t\right)^2 < \chi^2 < \frac{1}{2}\left(\sqrt{2k-1} + t\right)^2.$$

Таким чином, при обчисленні довірчого інтервалу при $k > 30$ потрібно вважати $\chi_1^2 = \frac{1}{2}\left(\sqrt{2k-1} - t\right)^2$, $\chi_2^2 = \frac{1}{2}\left(\sqrt{2k-1} + t\right)^2$, де $\Phi(t) = \gamma$.

3.5. Довірчий інтервал для ймовірності успіху в схемі Бернуллі і параметра λ розподілу Пуассона

Якщо розподіл генеральної сукупності не є нормальним, то в окремих випадках за вибірками великого обсягу можна побудувати довірчі інтервали для невід'ємних параметрів наближенно, використовуючи граничні теореми теорії ймовірностей, асимптотичні розподіли і оцінки.

Нехай в n незалежних випробуваннях подія A настала x разів. Знайдемо довірчий інтервал для ймовірності p настання в одному випробуванні. Відомо, що ефективною оцінкою успіху p в одному випробуванні є відносна частота $\overline{p} = \frac{x}{n}$. За теоремою Муавра-Лапласа відносна частота $\omega = \frac{x}{n}$ має асимптотично нормальний розподіл $N\left(p, \sqrt{\frac{pq}{n}}\right)$; $(q = 1 - p)$.

Розглянемо статистику $U = \dfrac{\frac{x}{n} - p}{\sqrt{\frac{pq}{n}}}$, яка має асимптотично нормальний розподіл $N(0,1)$ незалежно від p. Функцію цього розподілу позначимо:

$\Phi_0(x) = \frac{x}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$. При великих n:

$$P\left(\left|\frac{\frac{x}{n} - p}{\sqrt{\frac{pq}{n}}}\right| < u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha = \gamma, \ (\alpha = 1 - \gamma).$$

Звідси випливає, що з ймовірністю $1 - \alpha = \gamma$ виконується нерівність:

$$\frac{x}{n}-u_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}<p<\frac{x}{n}+u_{1-\frac{\alpha}{2}}\sqrt{\frac{pq}{n}}.$$

Замінимо значення p і q у цій нерівності їх оцінками $\overline{p}=\frac{x}{n}$ і $\overline{q}=1-\frac{x}{n}$. В результаті одержимо довірчий інтервал для ймовірності успіху в схемі Бернуллі: $\frac{x}{n}-u_{1-\frac{\alpha}{2}}\sqrt{\frac{x}{n^2}\left(1-\frac{x}{n}\right)}<p<\frac{x}{n}+u_{1-\frac{\alpha}{2}}\sqrt{\frac{x}{n^2}\left(1-\frac{x}{n}\right)}$.

Приклад 3.4. Із великої партії чипів одного типу було відібрано і перевірено 100 чипів. У 36 чипів було виявлено дефект (певний показник набував значень менших 10). 1) Знайти 95%-вий довірчий інтервал для частки бракованих чипів у всій партії. 2) Який мінімальний обсяг вибірки треба зробити, щоб із ймовірністю 0,95 можна було стверджувати, що частка бракованих чипів у всій партії відрізняється від частоти появи їх у вибірці не більше ніж на 1%.

Розв'язання. 1) Оцінка частки бракованих чипів: $\overline{p}=\frac{36}{100}$. За умовою

$1-\alpha=\gamma=0{,}95\Rightarrow\alpha=0{,}05;\ \frac{\alpha}{2}=0{,}025;\ 1-\frac{\alpha}{2}=0{,}975$. Отже $u_{1-\frac{\alpha}{2}}=u_{0,975}$.

Існує спеціальна таблиця для значень функції Φ_0. Але можна скористатися стандартною таблицею значень функції Лапласа, оскільки $\Phi_0(x)=0{,}5+\Phi(x)$. $\Phi(x)=-0{,}5+\Phi_0(x)=0{,}975-0{,}5=0{,}475$. За таблицею

$u_{0,975}=t_{0,475}=1{,}96$.

Довірчий інтервал для частки бракованих чипів:

$$0{,}36-1{,}96\sqrt{\frac{36}{10000}(1-0{,}36)}<p<0{,}36+1{,}96\sqrt{\frac{36}{10000}(1-0{,}36)}.$$

$0{,}266<p<0{,}454$.

2) Запишемо довірчий інтервал у вигляді: $\left|\frac{x}{n}-p\right|<t_{1-\frac{\alpha}{2}}\sqrt{\frac{x}{n^2}\left(1-\frac{x}{n}\right)}$.

За умовою $\left|\frac{x}{n}-p\right|\le\delta\ \ (\delta=0{,}01)$. Для визначення n

$$t_{1-\frac{\alpha}{2}}\sqrt{\frac{x}{n^2}\left(1-\frac{x}{n}\right)}\le\delta\Leftrightarrow t_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{p}\left(1-\overline{p}\right)}{n^2}}\le\delta\quad\left(\overline{p}=\frac{x}{n}\right).$$

Звідки, $n\ge\frac{t^2_{1-\frac{\alpha}{2}}}{\delta^2}\overline{p}\left(1-\overline{p}\right)$. Підставляючи відповідні значення, отримаємо:

$n\ge\frac{1{,}96^2}{\left(0{,}01\right)^2}0{,}36\left(1-0{,}36\right)=8851{,}0464$. Мінімальний обсяг вибірки

$n=8852$.►

Приклад 3.5. Політолог вирішив оцінити частку виборців, які проголосують за кандидатів правих сил на наступних виборах. Побудувати 90%-ий довірчий інтервал для цього прогнозу для оцінки із граничною помилкою вибірки $\pm 0{,}04$. Який обсяг вибірки знадобиться політологу для опитування виборців?

Розв'язання. Скористаємось формулою $n\ge\frac{t^2_{1-\frac{\alpha}{2}}}{\delta^2}\overline{p}\left(1-\overline{p}\right)\quad\left(\delta=\pm 0{,}04\right)$.

За умовою $1-\alpha=\gamma=0{,}9\Rightarrow\alpha=0{,}1;\ \frac{\alpha}{2}=0{,}05;\ 1-\frac{\alpha}{2}=0{,}95;\ u_{1-\frac{\alpha}{2}}=u_{0{,}95}$.

$\Phi\left(x\right)=-0{,}5+\Phi_0\left(x\right)=0{,}95-0{,}5=0{,}45$. За таблицею $u_{0{,}95}=t_{0{,}45}=1{,}645$.

Тоді, $n\approx\frac{1{,}645^2}{\left(0{,}04\right)^2}\overline{p}\left(1-\overline{p}\right)$. Оскільки немає інформації про значення $\overline{p}$,

розглянемо *n* як функцію від $\overline{p}$ і знайдемо її максимум:

$\left(n\left(\overline{p}\right)\right)'=\frac{1{,}645^2}{\left(0{,}04\right)^2}\left(1-\overline{p}-\overline{p}\right)=1691{,}27\left(1-2\overline{p}\right)=0\Rightarrow\overline{p}=0{,}5$. Отже,

$n_{max}=1691{,}27\cdot 0{,}5\cdot 0{,}5=422{,}82$. Отже, обсяг вибірки виборців $n=423$.

Довірчий інтервал: $0{,}5-0{,}04<p<0{,}5+0{,}04\Leftrightarrow 0{,}46<p<0{,}54$.►

Приклад 3.6. На кожній з 36 АТС міста у період з двох до трьох годин зафіксовано в середньому 2 виклики. Число викликів для кожної АТС має розподіл Пуассона з одним і тим самим параметром λ . Знайти довірчий інтервал для λ з надійністю $\gamma = 0,9$.

Розв'язання. Нехай маємо вибірку $x_1, ..., x_n$ з генеральної сукупності, яка має розподіл Пуассона. Тоді, довірчий інтервал має вигляд:

$$\overline{x}_B - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{x}_B}{n}} < \lambda < \overline{x}_B + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\overline{x}_B}{n}}.$$

За умовою, $1-\alpha = \gamma = 0,9 \Rightarrow \alpha = 0,1;$ $\frac{\alpha}{2} = 0,05;$ $1-\frac{\alpha}{2} = 0,95;$ $u_{1-\frac{\alpha}{2}} = u_{0,95}$.

$u_{0,95} \approx 1,65$. Підставимо знайдені значення:

$$2 - 1,65\sqrt{\frac{2}{36}} < \lambda < 2 + 1,65\sqrt{\frac{2}{36}} \Leftrightarrow 1,61 < \lambda < 2,39 .$$►

Контрольні питання

1. Що таке оцінка параметру розподілу?
2. Для чого використовують нерівність Рао-Крамера-Фреше?
3. Як побудувати довірчий інтервал?