*Міжнародний науково-технічний журнал*
***«ВИМІРЮВАЛЬНА ТА ОБЧИСЛЮВАЛЬНА ТЕХНІКА В ТЕХНОЛОГІЧНИХ ПРОЦЕСАХ»***
***ISSN 2219-9365***

УДК 681.12:519:168

О. О. ДРЮЧИН,
О. М. ВОЗНЯК,
В. М. СЕВАСТЬЯНОВ,
В. С. МАНЬКОВСЬКА
Вінницький національний технічний університет

# ВИЗНАЧЕННЯ ОПТИМАЛЬНОГО ОБ'ЄМУ ВИБІРКИ ПІД ЧАС ПРОВЕДЕННЯ ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕНЬ

*В роботі розглянуто варіанти оцінок вибірок довільного обсягу. Для них вирішується питання визначення мінімального обсягу вибірки, щоб отримана оцінка параметра розподілу або початкових умов задовольняла встановлені вимоги до проведення вимірювань з необхідною точністю.*
*Ключові слова: вимірювання, вибірка, обсяг вибірки, похибка, помилка, нормальний закон розподілу, ймовірність.*

O. DRUCHUN,
O.VOZNYAK,
O. SEVASTYANOV,
V. MANKOVSKA
Vinnytsia National Technical University

# DETERMINATION OF OPTIMAL SAMPLE SIZE DURING EXPERIMENTAL RESEARCH

*The variants of estimations of selections of arbitrary volume are considered in work. The question of determination of minimum sample size decides for them, that the got estimation of parameter of distributing or initial conditions satisfied the set requirements to conducting of measurings with necessary exactness. The use of a particular technique depends on the a priori information about the parameters of the distribution of random errors. If such information is not available, extensive experimental studies should be carried out. The article analyzed the ways of determining the optimal volume of tests in conducting multiple measurements of physical quantities and obtained analytical expressions for various cases concerning a priori information about the parameters of the distribution of random errors of measurement results. The use of the function of most plausibility is appropriate for a sufficiently large sample size. With a known confidence rating, with a measure of reliability, the sample size is determined with the help of t-distribution of the Student. To determine the membership of a sample to the general population, the required number of multiple measurements is determined at a known boundary, which estimates the errors of the 1st and 2nd kind. The ratios given in this article make it possible to reduce the cost of conducting research, which in many cases requires significant financial costs.*
*Keywords: measurement, sampling, sample size, measurement error, error, normal distribution law, probability.*

ВСТУП. Проведення вимірювального експерименту завершується опрацюванням результатів спостережень для визначення результату вимірювань – кінцевої мети вимірювань. В процесі цього вирішують дві задачі: першу – знаходять найкращу для вибраних методик, засобів, умов та отриманих даних оцінку значення вимірювальної величини і другу – оцінюють характеристики точності вимірювань. Результат вимірювання є повноцінним за умови, що він супроводжується оцінкою його точності, до якої відносяться багато важливих даних, зокрема – кількість спостережень.

Обсяг опрацювання залежить від різновиду вимірювання, кількості експериментальних даних, вимог до точності вимірювання, апріорної інформації про систематичні та випадкові похибки вимірювання тощо. Лише при прямих разових вимірюваннях отримані результати вимірювання. В інших вимірюваннях опрацювання може здійснюватись за стандартними методиками. Незалежно від виду розподілу випадкових похибок якість оцінювального результату покращується зі збільшенням результатів спостережень.

В роботі розглянуто варіанти вибірок довільного обсягу. Для них вирішується питання визначення мінімального обсягу вибірки, щоб отримана оцінка параметра розподілу або початкових умов задовільна встановленим вимогам до проведення вимірювань з необхідною точністю.

Перевагу варто надавати вибіркам із більшим обсягом, але зазвичай більший обсяг вибірки потребує й більших витрат для її одержання та обробки. Тому варто визначити, яким має бути мінімальний обсяг вибірки, щоб $P(|\overline{X}-X|<\varepsilon)=P_{дов}$, тобто ймовірність того, що результат вимірювань буде відрізнятись від дійсного значення на визначену величину $\varepsilon$, буде відповідати встановленій довірчій ймовірності.

МЕТА РОБОТИ. Ця робота присвячена аналізу способів визначення оптимального об'єму вибірки при проведенні незалежних вимірювань фізичної величини з дотриманням певних вимог.

***ISSN 2219-9365***

Використання функції найбільшої правдоподібності: Нехай дана вибірка ( $x_1 ... x_n$ ) об'єму $n$ з генеральної сукупності переведено розділеної вершини $x$. Функцією найбільшої правдоподібності називають функцію параметра $\gamma$ , що визначається співвідношенням [1]

$$L(x_1,\ldots,x_n;\gamma) = f(x_1,\gamma)f(x_2,\gamma)\ldots f(x_n,\gamma)\,, \tag{1}$$

де $\gamma$ – параметр розподілу.

В якості оцінки параметру $\gamma$ береться значення при якому функція правдоподібності досягає свого максимуму. Це означає. Це значення $\bar{\gamma}$ є функцією $x_1,\ldots,x_n$

$$\bar{\gamma} = \Gamma(x_1,\ldots,x_n) \tag{2}$$

Відповідна функція вибірки $\Gamma(X_1,\ldots,X_n)$ називається оцінкою найбільшої правдоподібності $\gamma$. Параметр $\bar{\gamma}$ знаходять рішенням відносно $\gamma$ рівнянням

$$\frac{\partial L}{\partial \gamma} = 0\,,$$

або

$$\frac{1}{L}\frac{\partial L}{\partial \gamma} = \frac{\partial \ln L}{\partial \gamma} = 0\,. \tag{3}$$

Якщо розподіл $X$ залежить від параметрів $l$ параметрів, то найбільш правдоподібну оцінку $\bar{\gamma}$ знаходять рішенням системи рівнянь

$$\frac{\partial L}{\partial \gamma_i} = 0 \qquad (i = 1,\ldots,l)\,,$$

Або

$$\frac{\partial \ln L}{\partial \gamma_i} = 0 \qquad (i = 1,\ldots,l)\,.$$

Для прикладу розглянемо оцінку невідомих параметрів $a$ і $\sigma$ нормального розподілу, виходячи з вибірки ( $x_1 ... x_n$ ) об'єму $n$ .

Функція правдоподібності [3]

$$L(x_1,\ldots,x_n;a,\sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k - a)^2\right]. \tag{4}$$

Відповідно

$$\ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_n - a)^2\,. \tag{5}$$

Запишемо систему рівнянь для визначення $a$ і $\sigma^2$

$$\begin{cases} \dfrac{\partial \ln L}{\partial a} = \dfrac{1}{\sigma^2}\sum\limits_{k=1}^{n}(x_n - a) = 0, \\ \dfrac{\partial \ln L}{\partial\left(\sigma^2\right)} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^2}\sum\limits_{k=1}^{n}(x_n - a)^2 = 0. \end{cases} \tag{6}$$

Шляхом вирішення системи знаходимо оцінку найбільшої правдоподібності параметрів $a$ і $\sigma^2$

$$\bar{a} = \frac{1}{n}\sum_{k=1}^{n} x_k = \bar{x}, \tag{7}$$

$$\overline{\sigma^2} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2. \tag{8}$$

Виходячи з аналогічних міркувань можливо оцінити оптимальний об'єм вибори результатів вимірювання фізичних величин при поставлених умовах до точності проведення експерименту.

Визначення об'єму $n$ при визначенні довірливої оцінки $a$ з невідомим $\sigma$.

Оцінка заснована на факті, що величина $\frac{\overline{\mathrm{X}} - a}{S}\sqrt{n}$ задовольняє $t$ - розподілу з $m = n-1$ ступенями свободи.

По таблицям $t$ - розподілу Стьюдента і заданому $a$ можливо знайти квантіль розподілу число $t_{\alpha,n-1}$ для якого справедливий вираз [1]

$$P(-t_{\alpha,n-1} < \frac{\overline{\mathrm{X}} - a}{S}\sqrt{n} < t_{\alpha,n-1}) = 1-\alpha, \tag{9}$$

або

$$P(\overline{\mathrm{X}} - \frac{s}{\sqrt{n}}t_{\alpha,n-1} < a < \overline{\mathrm{X}} + \frac{s}{\sqrt{n}}t_{\alpha,n-1}) = 1-\alpha. \tag{10}$$

Таким чином випадковий інтервал $(\overline{\mathrm{X}} - \frac{s}{\sqrt{n}}t_{\alpha,n-1}; \overline{\mathrm{X}} + \frac{s}{\sqrt{n}}t_{\alpha,n-1})$ є довірча оцінка $a$ з мірою надійності $p = 1-\alpha$ по якій можливо визначити оптимальний об'єм вибірки $n$ результатів вимірювання.

Визначення об'єму вибірки $n$ при визначенні довірливої оцінки $\sigma$ з невідомим $a$.

Оцінка базується на факті, що при заданих умовах величина $\frac{(n-1)s^2}{\sigma^2}$ задовольняє $\mathrm{X}^2$ - розподіл з $n-1$ ступенями свободи. З таблиці $\mathrm{X}^2$ - розподілу по заданим $a$ і $m = n-1$ ступеням свободи визначають два числа $c_1$ і $c_2$, так, що

$$P(\mathrm{X}^2 < c_1) = 1 - P(\mathrm{X}^2 > c_1) = \frac{\alpha}{2}, \tag{11}$$

і

$$P(\mathrm{X}^2 > c_2) = \frac{\alpha}{2}. \tag{12}$$

Числа $c_1 = \mathrm{X}^2{}_{1-\frac{\alpha}{2}}$, $c_2 = \mathrm{X}^2{}_{\frac{\alpha}{2}}$ визначенні так, що

$$P(c_1 < \frac{(n-1)s^2}{\sigma^2} < c_2) = 1-\alpha, \tag{13}$$

або

$$P(\frac{(n-1)s^2}{c_2} < \sigma^2 < \frac{(n-1)s^2}{c_1}) = 1-\alpha. \tag{14}$$

Таким чином $(\frac{(n-1)s^2}{c_2}; \frac{(n-1)s^2}{c_1})$ є довільною оцінкою $\sigma^2$ з мірою надійності $1-\alpha$. Виходячи з вищенаведених міркувань можливо визначити об'єм вибірки $n$ результатів незалежних вимірювань фізичної величини, за умови, що СКВ результатів не перебільшує наперед заданого значення.

Визначення об'єму вибірки n результатів незалежних вимірювань розподілених за нормальним законом

Нехай при проведенні незалежних вимірювань похибка розподілена за нормальними законами. СКВ задане і накладаються обмеження, щоб хоча б для одного з вимірювань похибки не перебільшувала $\pm\Delta$, з ймовірністю $P_3$. Ймовірність того, що при одному вимірюванні похибка не перебільшує $\pm\Delta$, дорівнює

$$P = P[\left|\Delta_{1,2}\right| < \Delta] = \Phi(\frac{\Delta}{\sigma}),$$

де $\Phi(2)$ – функція Лапласа.

Ймовірність того, що при $n$ незалежних вимірюваннях ні одне з них не забезпечить похибки меншої $\pm\Delta$, дорівнює

$$(1-P)^n = [\Phi(\frac{\Delta}{\sigma})]^n, \quad (15)$$

що знаходять, скориставшись таблицями функції Лапласа.

З іншого боку, ця сама ймовірність повинна бути не більша ніж $P_3$.

$$1 - P_3 = \alpha$$

Тому можливо записати нерівність

$$[\Phi(\frac{\Delta}{\sigma})]^n \le \alpha. \quad (16)$$

Звідки

$$n \ge \frac{\lg\alpha}{\lg[\Phi(\frac{\Delta}{\sigma})]} = \lg[\alpha - \Phi(\frac{\Delta}{\sigma})]. \quad (17)$$

Тоді оптимальний об’єм вибірки результатів незалежних вимірювань буде визначатись як найменше ціле число, що задовольняє нерівність.

Розглянемо інший випадок коли необхідно визначити мінімальний обсяг вибірки, щоб

$$P(\left|\overline{a} - a\right| < \varepsilon) = P_{дов}, \quad (18)$$

де $a$ – середнє арифметичне.

Середньоквадратичне значення $\sigma$ для генеральної сукупності відоме і визначене при попередніх дослідженнях. За результатами поточних вимірювань $x_i$ знаходять середнє значення $\overline{x}$, яке має нормальний закон розподілу, незалежно від вихідного розподілу генеральної сукупності.

Розглянемо випадок для нормованої випадкової величини, коли $a = \mathrm{MX}$ і $\overline{a} = \overline{x}$, тоді [2]

$$P(\left|\frac{\overline{x} - \mathrm{MX}}{\sigma\overline{x}}\right| < z_{1-\frac{1}{\alpha}}) = 1 - \alpha. \quad (19)$$

Зіставивши попередні вирази, можна записати, що припустиме відхилення середнього від математичного сподівання з ймовірністю $1-\alpha$ не перевищує $\varepsilon = \sigma \cdot z_{1=\frac{\alpha}{2}}$.

Скористаємось тим, що дисперсія середнього в $n$ разів менша за вихідну дисперсію, запишемо

$$\varepsilon = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}. \quad (20)$$

Тоді мінімальний обсяг вибірки, тобто мінімальна кількість спостережень $n_{\min}$, з якої отримане середнє буде відрізнятись від істинного значення не більше ніж на $\varepsilon$ з ймовірністю $1-\alpha$

$$n_{\min} = \frac{z_{1-\frac{\alpha}{2}}^{2}\sigma^2}{\varepsilon^2}. \quad (21)$$

Визначення об’єму вибірки при перевірці гіпотези про належність вибірки до генеральної сукупності.

***ISSN 2219-9365***

Припустимо, що на вхід приладу може бути подано одне з двох значень величини $x_0$ і $x_1$. При багаторазовому вимірюванні вихідної величини отримаємо значення $x_i = (i = 1,\ldots,n)$, та обчислимо середнє $\overline{x}$. Необхідно визначити, чи є $\overline{x}$ оцінкою випадкової величини із центрами $x_0$ і $x_1$. Висувається гіпотеза Но $: \overline{x} \in x_0$ при альтернативній гіпотезі $Hi : \overline{x} \in x_i$. [2]

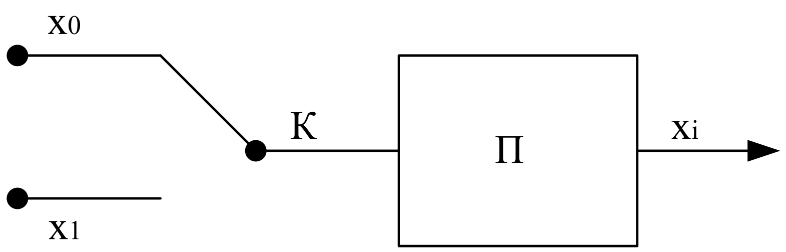


**Рис. 1. Перевірка гіпотези про належність вибірки до генеральної сукупності**

Ймовірність виникнення помилки першого роду

$$\alpha = P[(\overline{x} - x_0) > \varepsilon], \tag{22}$$

де $\varepsilon = z_{1-\alpha}\sigma_{\overline{x}}$.

Ймовірність виникнення помилки другого роду

$$\beta = P[(\overline{x} - x_1) < \varepsilon]. \tag{23}$$

Виходячи з оптимальної процедури перевірки гіпотез, необхідно визначити мінімальний обсяг випробувань, який забезпечує помилки першого й другого роду, що не перевищують заданих значень.

Для гіпотези Но можна записати

$$P[(\overline{x} \le \frac{\varepsilon}{x_0}] = 1 - \alpha, \tag{24}$$

де $\varepsilon$ – фіксоване значення, що забезпечує одночасно значення $\alpha$ і $\beta$, менші від заданих.

Тоді $\dfrac{\varepsilon - x_0}{\dfrac{\sigma}{\sqrt{n}}} = z_{1-\alpha}$.

Для гіпотези $Hi$ має виконуватись співвідношення

$$P(\overline{x} \le \frac{\varepsilon}{x_i}) = \beta . \tag{25}$$

В цьому разі $\dfrac{\varepsilon - x_1}{\dfrac{\sigma}{\sqrt{n}}} = z_\beta$

Після перетворень маємо

$$\sqrt{n}(x_1 - x_0) = z_{1-2}\sigma - z_\beta\sigma \tag{26}$$

Звідки кількість випробувань

$$n_{\min} = [\frac{\sigma(z_{1-\alpha} - z_\beta)}{x_1 - x_0}]^2 . \tag{27}$$

Якщо границя $\varepsilon$ задана, обчислену кількість випробувань $n$ округлюють до цілого числа в більшу сторону.

ВИСНОВКИ. Використання конкретної методики залежить від апріорної інформації про параметри розподілу випадкових похибок. Якщо така інформація відсутня, то необхідно виконати широкі за обсягом експериментальні дослідження. В статті було проведено аналіз шляхів визначення оптимального об'єму випробувань при проведенні багаторазових вимірювань фізичних величин і отримані аналітичні вирази для різних випадків стосовно апріорної інформації про параметри розподілу випадкових похибок результатів вимірювань. Використання функції найбільшої правдоподібності доцільно при достатньо великому об'ємі вибірки. При відомій довірчій оцінці $\alpha$ з мірою надійності $\rho$ обсяг вибірки визначається з допомогою t - розподілу Стьюдента. Для встановлення належності вибірки до генеральної сукупності необхідну кількість багаторазових вимірювань визначають при відомій границі $\varepsilon$, яка оцінює помилки I та II роду.

Наведені в статті співвідношення дають можливість зменшити затрати на проведення досліджень, для яких в багатьох випадках потребуються значні фінансові витрати.

## Література

1. Бронштейн И.Н. Справочник по математике / Бронштейн И.Н., Семендяев К.А. – М. : Наука. Главная редакция физико-математической литературы, 1981. – 720 с.

2. Володарський Є.Т. Статистична обробка даних : навчальний посібник / Володарський Є.Т., Кошева Л.О. – К. : НАУ, 2008. – 308 с.

3. Основи метрології та вимірювальної техніки : підручник : у 2 т. Т. 1. Основи метрології / Дороժовець М., Мотало В., Ковальчик А. та ін. ; за ред. Б. Стадник. – Львів : Видавництво Національного університету «Львівська політехніка», 2005. – 532 с.

## References

1. Bronshteyn I.N. Spravochnik po matematike / Bronshteyn I.N., Semendyaev K.A. – M. : Nauka. Glavnaya redaktsiya fiziko-matematicheskoy literaturyi, 1981. – 720 s.

2. Volodarskyi Ye.T. Statystychna obrobka danykh : navchalnyi posibnyk / Volodarskyi Ye.T., Kosheva L.O. – K. : NAU, 2008. – 308 s.

3. Osnovy metrolohii ta vymiriuvalnoi tekhniky : pidruchnyk : u 2 t. T. 1. Osnovy metrolohii / Dorozhovets M., Motalo V., Kovalchyk A. ta in. ; za red. B. Stadnyk. – Lviv : Vydavnytstvo Natsionalnoho universytetu «Lvivska politekhnika», 2005. – 532 s.